

VISIT...

LANZAROTE
Caliente.COM

Perin, James C. Ware,

BASIC MATHEMATICS IN ELECTRICAL COMMUNICATIONS

A TREATISE
ON
CONIC SECTION CURVES, EXPONENTIALS
ALTERNATING CURRENT, ELECTRICAL OSCILLATIONS,
TRANSIENTS AND HYPERBOLIC FUNCTIONS

NATURE SEEMS TO BE PLEASED WITH SIMPLE THINGS
NEWTON

ALL GREAT TRUTHS ARE SIMPLE
ANON.

Upper Montclair
New Jersey
September 1957

This pre-publication copy of "Basic Mathematics in Electrical Communications" is presented to the Library of the Physics Department of Cornell University in grateful appreciation of the excellent instruction and inspiration of E. Merritt, F. Bedell and F. K. Richtmyer and the pleasant and helpful associations with Professors C. C. Bidwell, H. Howe, R. C. Gibbs, C. C. Murdock and G. E. Grantham.

J. Owen Perrine

J. Owen Perrine
B.A., Iowa, 1909
Ph.D., Cornell, 1921
Captain, Signal Corps, 1918-19

Table of Contents

	<i>Page</i>
Introduction.....	2
Chapter 1. The Conic Sections.....	7
Chapter 2. Circle, Sine, Cosine, Areas, Simple Harmonic Motion.....	21
Chapter 3. Ellipse, Simple Harmonic Motion.....	32
Chapter 4. Parabola, Simple Harmonic Motion, Mirror.....	40
Chapter 5. Hyperbola, Area.....	60
Chapter 6. Hyperbolic Sinh, Cosh, Tanh.....	69
Chapter 7. Fundamental Considerations of Conic Sections.....	87
Chapter 8. Natural Law of Geometric Retrogression, Exponential Equations, Attenuation, Decibels, Alternating Current Decay.....	96
Chapter 9. ^{Part 2} Resistance, Inductance, Capacitance, Oscillating Circuits.....	139
Chapter 10. Kelvins Equation for Electrical Oscillations.....	162
Chapter 11. Exponential Equations and Transient Curves in R , L and C Circuits.....	175
Chapter 12. Exponentials, Vectors Series Expansion of Sine, Cosine, Sinh, Cosh.....	225
Chapter 13. Applications of Sinh and Cosh.....	244
Chapter 14. Infinitely Long Electrical Circuits, Use of Sinh and Cosh.....	248
Epilogue.....	259

Preface

It is believed that the book may be of value and help to several groups of technically minded and trained people. These groups are:

1. The officers and men in all branches of the Armed Forces, whose job relates to mathematics, electronics and electrical communication. In this category are not only the men in training during their term of service in the Armed Forces but also those men in the Reserve Officers Training Corps program in colleges and universities.

2. The junior and senior college student who is taking courses in physics, mathematics and electrical engineering.

3. The young instructor in college who is seeking to enlarge and integrate his "know how" of classroom subject matter and to acquire a feeling of mastery of the fundamentals in his field of endeavor.

4. The many men who are allied with technical endeavors but who did not get to go to college. With great zeal and commendable enterprise, these men go to night school and do much home reading and study. It is rather difficult for these men to teach themselves by use of the conventional textbook. It is believed that they could learn and grasp the principles and doctrines presented in this book without the help of a teacher.

Learning is a slow and continuing process. It takes time to acquire knowledge. Learning is not a "one shot" affair. New ideas are met for the first time, and then need to be meditated on many times. The ideas and concepts to be learned need to be expressed with a wide variety of different words and points of view. This teaching doctrine does not mean verbosity and redundancy. Repetitions and several reviews are necessary. Brevity may be the soul of wit, but not of learning and understanding. Ideas require a long time to sink in. Hence, there are more words, drawings, curves and tables per idea than ordinarily found in technical treatises.

TWO REMARKABLE BOOKS

There are two mathematical books that have been a delight to this writer. "Calculus Made Easy" published by Dr. S. P. Thompson in 1914 and "Exponentials Made Easy" published by Gheury de Bray in 1928 are those books. These two men, particularly the former, were outspoken and harsh "debunkers" of the conventional. Both wrote with a trenchant pen. They are most helpful and stimulating books and can be read and studied with great profit. They are to be strongly recommended for all who wish to get a thorough understanding.

In the Introduction of his book, De Bray writes "Some time ago, the author came across a truly delightful little book, "Calculus Made Easy." Although he was supposed to know about the things explained, he found great delight in reading it. In so doing, he re-learned a number of things he had forgotten and learned for the first time several others he had not chanced to meet before. This newcomer has no pretension to equal its elder but it is setting forth with a desire to be worthy of kinship and it certainly could not choose a better example to emulate."

In the same vein as de Bray wrote of Dr. Thompson's book "Calculus Made Easy," so this writer, neither presumes to compare the present book with the two books named nor makes any pretension that it approaches its elders. This book is set forth with the desire and hope to be worthy of remote kinship. It could not select better examples to emulate.

ACKNOWLEDGEMENTS

The writer gratefully acknowledges the substantial help, wise counsel and kind encouragement of a number of old and new friends and associates in the preparation of this book. He is much indebted to Professor H. M. Turner of the Electrical Engineering Department of Yale University. Mr. Davis C. Greene an engineer of Upper Montclair, N. J. rendered most skillful assistance in making and cutting the cones for the conic sections. Mr. F. X. Kaminski of the Instructor Training Branch at Fort Monmouth also rendered skillful assistance in calculating data and making of curves. He particularly wishes to express his thanks to Mr. Jacobus Pieter Noteboom, a civilian instructor in the Signal Corps School at Fort Monmouth, New Jersey, for his important contributions. Mr. Walter C. Smith, a civilian member of the Signal Corps Board at Fort Monmouth and formerly of the Engineering Department of the Michigan Bell Telephone Company, not only made important contributions but also, took part in many discussions in the effort to crystallize and validate basic approach, thinking and conclusions.

Introduction

Engineering always has been concerned with efficiency. It has ever been the aim to get a high ratio of output to input. Many different qualitative and quantitative factors are involved in the functioning of engineering devices. Energy transformations must always be carefully considered. Engineers diligently seek to understand and control the various factors. They ever have had the objective of achieving greater efficiency.

It is significantly true that nature seems imperiously to decree that some factor among those involved shall decrease and that it shall decrease in geometric retrogression fashion. This decree is a natural law. This natural law says that the ratio between two successive values of the decreasing factor is constant. The ratio is of course numerically less than unity. The nearer the ratio is to unity, the smaller the rate of decrease and loss and the greater the efficiency.

The basic truth of geometric retrogression can be expressed nicely by a simple exponential equation. The equation relates the ratio, " r ," and the two variables, A and B .

$$A = A_0 r^B$$

If B is 1, then $A = .8A_0$. If B is 2, then $A = A_0 .8^2 = .64A_0$. If B is 3, then $A = A_0 .8^3 = .512A_0$. If $B = 10$, then $A = A_0 .8^{10} = .1073A_0$.

It is evident that the independent variable, B , plays the role of an exponent of the ratio. The exponent B is the logarithm and the base of the logarithmic system is $.8$. The equation is a simple one and nicely expresses the basic truth. It is called an exponential equation.

In the last of the above example, $A = A_0 .8^{10}$, one would not be so foolish as to multiply $.8$ by itself 10 times to get the answer. One could get the answer by that simple process, to be sure, and the basic truth of geometric retrogression would be well exemplified. However, it would be equally simple, and much easier as a matter of sheer calculation, to invoke simple logarithmic principles to get the answer. It would be better judgment to use logarithms to the base 10 or those to the base, 2.7128. This latter base is conventionally designated as e , the Greek letter epsilon. Fortunately, logarithmic tables to these two bases have been calculated by others in the past and tables thereof are happily available.

The problem, then, is to transform the simple exponential equation $A = A_0 .8^B$ into an equally simple logarithmic and workable form by use of the base 10 or the base 2.7128, whichever system one prefers. There is no particular advantage intrinsically embodied in either system.

Since $.8 = \frac{1}{1.25}$, one can readily write

$$A = A_0 \left(\frac{1}{1.25} \right)^B = A_0 \left(\frac{1}{1.25^B} \right) = A_0 \cdot 1.25^{-B}$$

The logarithm of 1.25 to the base 10 is .09691. Therefore,

$$A = A_0 \left(\frac{1}{10^{.09691}} \right)^B = A_0 10^{-.09691B}$$

So to get the answer easily to the last illustration previously mentioned, when $B = 10$

$$A = A_0 \left(\frac{1}{10^{.09691}} \right)^{10} = A_0 \frac{1}{10^{.09691 \cdot 10}} = A_0 \frac{1}{10^{.9691}}$$

Looking in the table of logarithms, $10^{.9691} = 9.311$. Now $\frac{1}{9.311} = .1073$. So $A = .1073A_0$. If the logarithmic

FOOTNOTE. In many places in this treatise, the dot is not only a decimal point in the conventional position but also the dot is centered on a line with the middle of the letters or numerals to indicate multiplication e.g. $4 \cdot 10 = 40$ $10^{2 \cdot 4} = 10^{12}$

base 2.7128 were chosen, then as before

$$A = A_0 \cdot .8^n = A_0 \frac{1}{1.25^n} = A_0 \left(\frac{1}{1.25} \right)^n = A_0 \cdot 1.25^{-n}$$

The log of 1.25 to the base 2.7128 (ϵ) is .22314. Therefore,

$$\begin{aligned} A &= A_0 \left(\frac{1}{\epsilon^{.22314}} \right)^n A_0 \epsilon^{-.22314 \cdot n} = A_0 \frac{1}{\epsilon^{.22314}} \\ &= A_0 \frac{1}{9.311} \\ &= .1073 A_0 \end{aligned}$$

In all cases where there is a constant ratio (less than unity) between successive terms of the dependent variable the exponent is positive in the basic exponential equation expressing the simple and basic truth of geometric retrogression. However, the simple equation can be written in terms of logarithms to the base 10 or the base 2.7128 (ϵ) to get answers more easily. To be sure, the negative sign is oftentimes used as per the conventional notation for exponents of reciprocals. Again, it is to be stressed however, that the exponent is really positive.

In some physical and engineering endeavors, there may be an increase of some factor. That increase will occur likely in geometric progression fashion. The constant ratio will be greater than unity and as before the sign of the exponent in the three equations is positive.

A good example of geometric progression is the increasing radius of the ever widening opening of a horn of circular cross section as a function of length. Such horns are used in public address loudspeaker systems, and are called exponential horns.

For a particular horn whose throat radius is one inch, whose mouth is 8 inches in radius, whose length is 60 inches, and whose ratio of growth is 1.035 (r) the basic exponential equation is

$$R = R_0 r^L = 8 = 1 \cdot 1.035^{60}$$

The logarithmic equations would be

$$8 = 1 \cdot 10^{.150 \cdot 60} = 1 \cdot \epsilon^{.0150 \cdot 2,3026 \cdot 60} = 1 \cdot \epsilon^{.0346 \cdot 60} = 1 \cdot \epsilon^{.20760} = 1 \cdot 10^{.9000}$$

The logarithmic constant to the base 10 is .0150, and to the base 2.7128 (ϵ) is .0346. Whatever system of logarithms is chosen for ease and convenience of calculation, the simple and basic fact is that the constant ratio of increasing radius per foot of length of horn is 1.035(r). In other words, the constant growth ratio is 3.5%. See page (108) and accompanying chart., Fig. 45.

In considering geometric progression and geometric retrogression phenomena, one frequently and fittingly encounters e^x and e^{-x} .

If one writes the simple equation, $Y = e^x$, understanding of its significance is provided by writing it, $y = 2.7183^x$. After all e is only a number of explicit numerical value, namely 2.7183. The e has no mysterious, esoteric nor ineffable nature.

$$Y = e^x = 2.7183^x; \text{ a geometric progression}$$

This simple equation states that the constant ratio between varying increasing values of y resulting from unitary arithmetical increases in x is 2.7183. The growth of y as per the constant ratio concept is in geometric progression. The constant ratio need not be 2.7183 but can be any value as prescribed by the intrinsic nature of the many phenomena which exists in our physical environment.

If one writes the simple equation, $Y = e^{-x}$, having a negative exponent, understanding and meaning is provided by following simple and valid mathematical maneuvers. It can be written $Y = \frac{1}{e^x}$. In turn, this can validly be written $Y = \left(\frac{1}{e} \right)^x$. Hereby, the exponent has been restored to a positive status. Since $\frac{1}{e}$ is only a number

with no magic connotation, namely $\frac{1}{2.7183}$ or .3679, one can write

$$Y = .3679^x; \text{ a geometric retrogression}$$

This simple mathematical maneuver of writing $Y = e^{-x}$ as $Y = \frac{1}{e^x} = \left(\frac{1}{e}\right)^x = .3679^x$ is most helpful and illuminating.

The equation so written states that the constant ratio between physical factors present in natural phenomena is less than unity and so identifies a geometric retrogression. In other words, the negative exponent encountered in so many mathematical equations relating to many simple natural phenomena immediately informs one that the constant geometric ratio will be less than unity. Here again as was the case of e and a constant ratio of 2.7183, the constant ratio .3679 is not a basic universal ratio decreed by nature. The ratio can be any value as prescribed by the intrinsic nature of phenomenon involved whether it relates heat, light, sound, electricity, mechanics, music, etc.

If one generalizes the former simple equation $Y = e^x$ by adding an increment constant c , one writes $Y = e^{cx}$. The c 's for different phenomenon vary

$$\text{If } c = .01; e^c = e^{.01} = 1.0101 \text{ and } Y = 1.0101^x$$

$$\text{If } c = 1.03; e^c = e^{1.03} = 2.8011 \text{ and } Y = 2.8011^x$$

$$\text{If } c = 2.01; e^c = e^{2.01} = 7.4633 \text{ and } Y = 7.4633^x$$

One generalizes again and writes

$$Y = e^{-Kx} = \frac{1}{e^{Kx}} = \left(\frac{1}{e^K}\right)^x$$

where K is a decrement constant.

$$\text{If } k = .01; \frac{1}{e^{.01}} = .9900 \text{ and } Y = .9900^x$$

$$\text{If } k = 1.03; \frac{1}{e^{1.03}} = .3570 \text{ and } Y = .3570^x$$

$$\text{If } k = 2.01; \frac{1}{e^{2.01}} = .1340 \text{ and } Y = .1340^x$$

If x has a value of 1 in all the above cases, the value of Y is obvious. If x has a value of 2 with the varying increment constants c , then one obtains by further use of e^x and e^{-x} tables the following values of Y in geometric progression.

$$Y = e^{.01 \cdot 2} = e^{.02} = 1.0101^2 = 1.0202$$

$$Y = e^{1.03 \cdot 2} = e^{2.06} = 2.8011^2 = 7.8460$$

$$Y = e^{2.01 \cdot 2} = e^{4.02} = 7.4633^2 = 55.701$$

For geometric retrogression involving a negative exponent and constant ratio less than unity the following calculations pertain for $x = 2$ and varying K 's.

$$Y = e^{-.01 \cdot 2} = e^{-.02} = .9900^2 = .9802$$

$$Y = e^{-1.03 \cdot 2} = e^{-2.06} = .3570^2 = .1274$$

$$Y = e^{-2.01 \cdot 2} = e^{-4.02} = .1340^2 = .0179$$

In the case of radium disintegration, the constant ratio of decreasing amounts per one year is .99957. The logarithmic decrement constant is .000438 and therefore $e^{-.000438} = .99957$. For a period of two years, one

could write

$$\begin{aligned} M &= M_0 e^{-.000435 \cdot 2} = M_0 e^{-.00087} = M_0 .99957^2 \\ &= M_0 .99904 \end{aligned}$$

For a period of 1580 years, the constant ratio of decrease is one half or .5; thus the expression of half life.

$$.5 = \frac{1}{2} = \frac{1}{e^{.6931}} = e^{-.6931}$$

For two periods of 1580 years, $\frac{1}{2}$ of $\frac{1}{2}$ or $\frac{1}{4}$ would be involved

$$.25 = \frac{1}{4} = \left(\frac{1}{2}\right)^2 = \frac{1}{e^{.6931 \cdot 2}} = e^{-1.3862}$$

$$M = M_0 e^{-.000435 \cdot 1580} = e^{-.6931} = .99957^{1580} = .5$$

Uranium of mass number 235 has a half life, a constant loss ratio of .5, of 7.07 times 10^8 years. This is over 200 million years, a very very long time. Polonium of mass number 212 has a half life of 3 times 10^{-7} seconds. This is three tenths of a microsecond, a very, very short time.

In all cases, where e^x , e^{-x} , e^{cx} and e^{-kx} are involved and the value of x is unity, it would be sensible to retain and use these formulas as such and find value of y directly from logarithmic tables.

In fact the simple principles of geometric progression and retrogression are not revealed by inspection of the formula per se. The formulas are directives as to how to find value of Y by use of logarithmic doctrines and logarithmic tables after one has an appreciation and clear understanding of the basic and simple truths involved.

The exponent is positive in all three forms of the equation whether there be geometric retrogression or progression. The logarithmic forms of the equation do not intrinsically embody the two simple truths. They merely provide directions to get an answer easily and conveniently by use of tables rather than by laborious multiplication.

In general then, decreases, decays or losses occur among the many factors encountered in engineering. The decrease may be thought of as a thinning out. The thinning out idea is well and conventionally expressed by the word *attenuation*; a geometric retrogression as contrasted to a geometric progression. This sixty-four dollar word need frighten no one. Although it has five syllables, it is a good word to convey a definite meaning.

In a broad and over-all way, nature's imperious law decrees that there shall be attenuation. As a result, there shall be an efficiency less than 100 percent. Attenuation is a grim reality and a universal fact which injects itself in almost every engineering problem. It demands one's first consideration.

Not only are exponential equations, especially those with negative exponents, regarded by many students of science and engineering as a bit frightening, but also the hyperbolic functions $\sinh$, $\cosh$, and $\tanh$ are oftentimes a "signal for undignified retreat."

Neither of these attitudes regarding these two great mathematical tools need stymie any one who will make a good try.

Actually, exponential equations per se and the hyperbolic functions per se can be rationalized and nicely understood by simple and basic algebraic, trigonometric, and geometric steps of analysis, as well as can the circular sine, cosine, and tangent. In this treatise the reader and student is guided not only to such a rationalization and understanding of these tools per se but also to their application to electrical and physical problems.

As one considers the broad gamut of problems of engineering in a fundamental way, attenuation is closely allied to exponentials and logarithms. In turn, exponentials can be basically allied numerically to hyperbolic functions. In turn, hyperbolic functions are, as their name implies, geometrically allied to the rectangular or equilateral hyperbola. The hyperbola is a conic section curve.

This book "takes a look" and logically considers in an integrated fashion the physical and engineering problems involving attenuation. As said before, attenuation, conics, hyperbolics and exponentials are intrinsically interrelated.

In many cases, satisfactory numerical answers in various problems could be found by simple arithmetic.

However, the calculations would be long and tedious. Logarithms and hyperbolic functions need not be asked to render assistance; but engineers want to be efficient in working problems of efficiency, so they seek the best "tools" available.

Engineers and scientists put the simple and basic facts and truths, which they have evolved regarding basic phenomena, into the symbolic and "shorthand" form of a simple mathematical equation. Then they seek the aid of formal mathematics as such and of mathematicians to get more explicit results. Very wisely, they seek to use the contributions, processes and results of mathematics per se as a "tool." Mathematics supplies the explicit relations and alliances of attenuation, conics, hyperbolics and exponentials. The final equations may appear to be elaborate, complex and even frightening, but they are not because they are basically founded on simple truths. Once thoroughly understanding and appreciating not only the basic truths of the problem at hand but also the equations and facts of alliance, the engineer used logarithmic and hyperbolic function tables, if they suit his purposes, to get quantitative answers.

It is one of the objectives of this book to show that mathematics is neither a "headache" nor "highbrow" stuff. The essentials are straight-forwardly, simple, rationally and logically considered and expounded.

It is well for the beginner and novice to pay respect and to admire the talented and experienced scientist, mathematician and engineer. However the beginner must be reminded that these men and women were once beginners also and possessed no secret talisman which vouchsafed success and achievement.

CHAPTER 1

The Conic Sections

The circle, ellipse, parabola and hyperbola are excellent "tools" with which the physical scientist and engineer explores and studies the quantitative and qualitative aspects of a wide variety of natural phenomena. In terms of the characteristics of these four curves, plus other curves too of course, the physicist, the astronomer, the architect and the mechanical, aeronautical, civil and electrical engineer can express and state, explicitly and lucidly, his understanding of observations, facts and truths presented him by nature. Furthermore, he can express thereby and share with others his thinking and conclusions regarding the principles and doctrines embodied in the basic evidence.

Curves and mathematical equations are a facile language in terms of which the solutions of many practical problems and the directives for many operations can be written.

The four curves above mentioned are called conic sections because they are outlined or created by the intersection, at various angles, of a plane and the surface of a cone. It is suggested that the reader right now take a good look at the accompanying photographs. During and after studying and digesting the discussion of the conic sections, many more looks might well be taken. In later pages (87, 90, 92) the conic section aspect of such of the four curves is analyzed and discussed in some measure of detail. It would be in order that the reader take a good preliminary look at photographs immediately following.

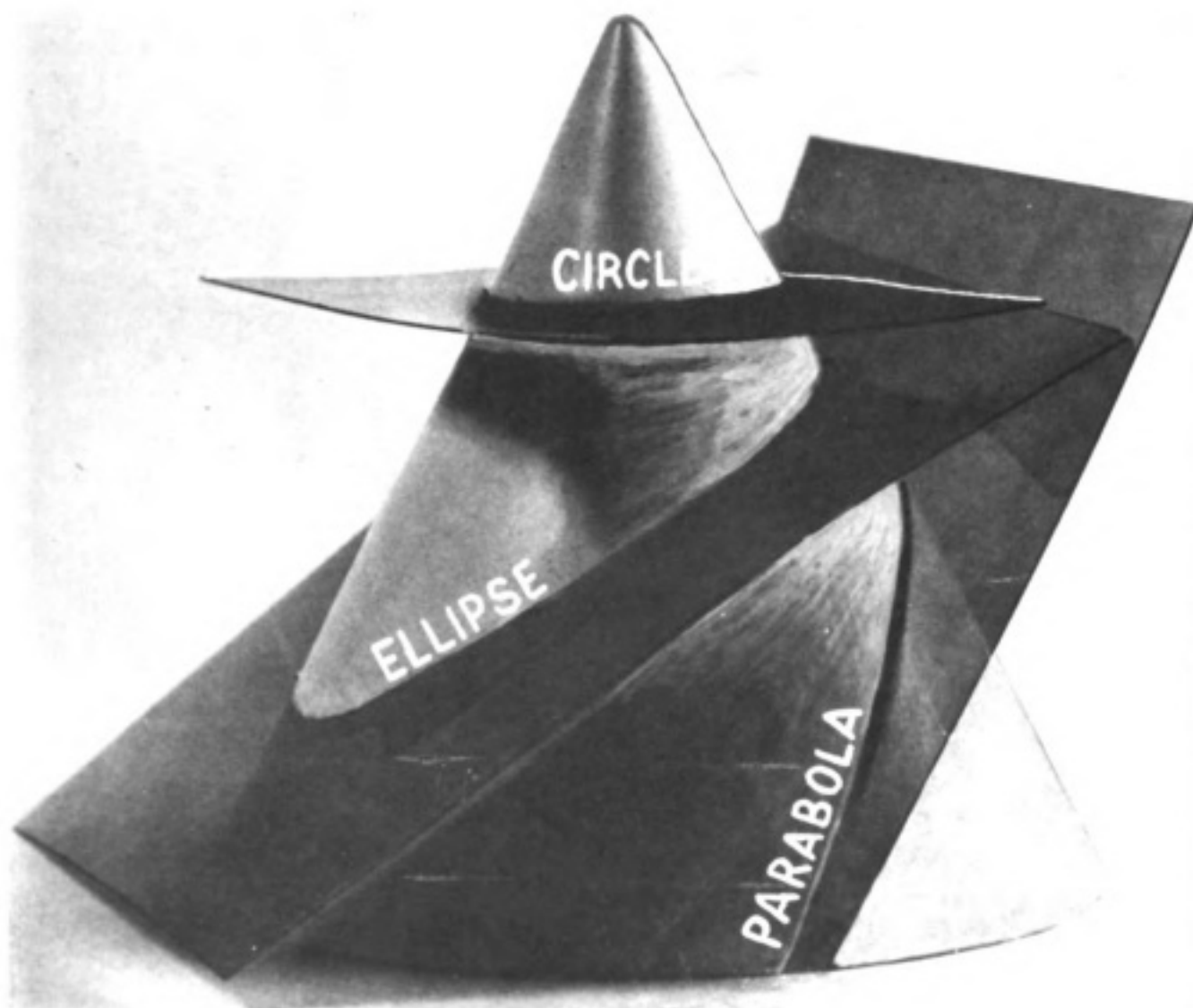


FIGURE 1

CIRCLE, ELLIPSE AND PARABOLA ARE PRODUCED BY THREE INTERSECTING PLANES. PLANE PARALLEL TO BASE PROVIDES CIRCLE; PLANE PARALLEL TO SLANT SURFACE PROVIDES PARABOLA; PLANE AT ANGLE BETWEEN THE OTHER TWO PLANES PROVIDES ELLIPSE.

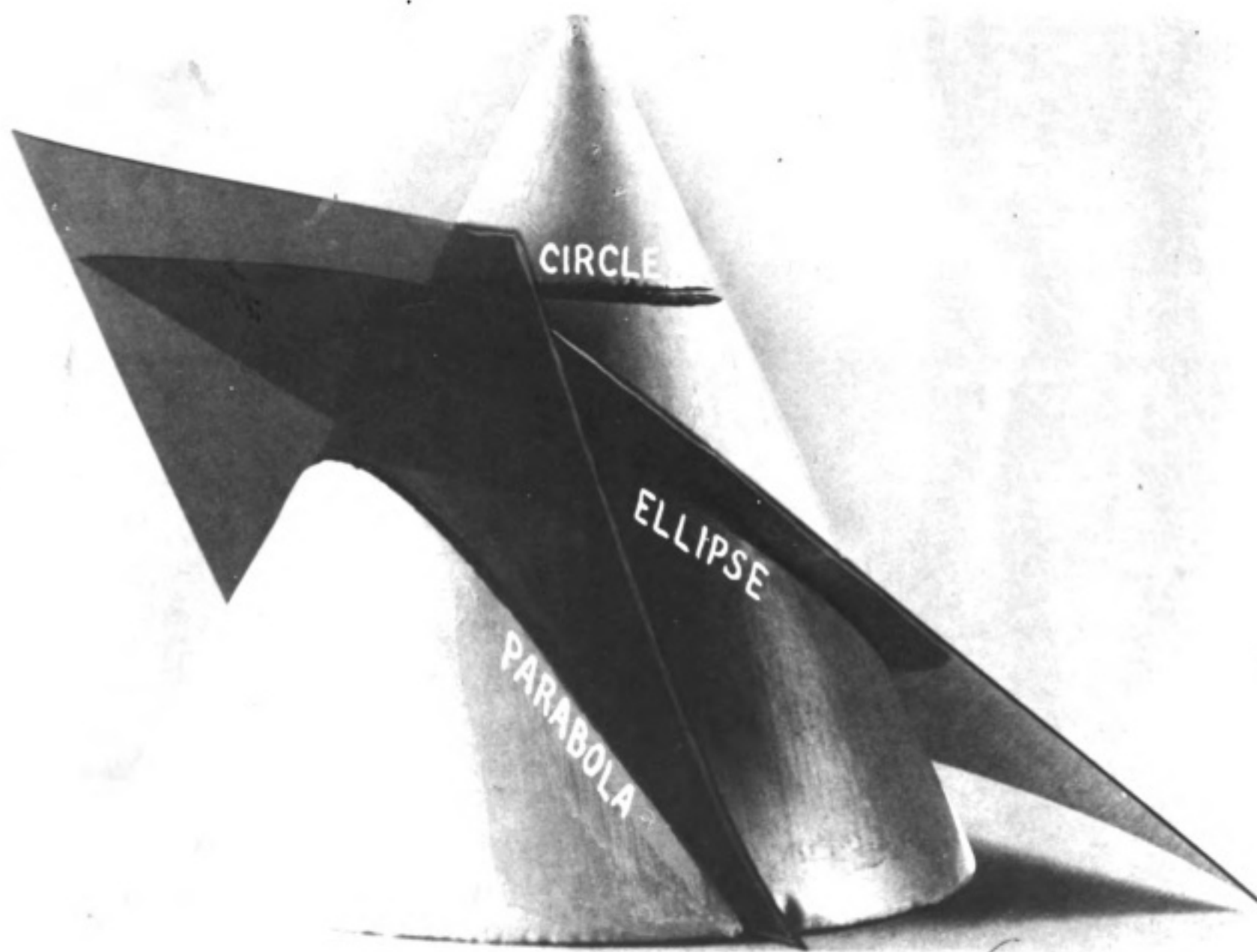


FIGURE 2

SAME MODEL AS FIG. 1, BUT FROM DIFFERENT CAMERA ANGLE.

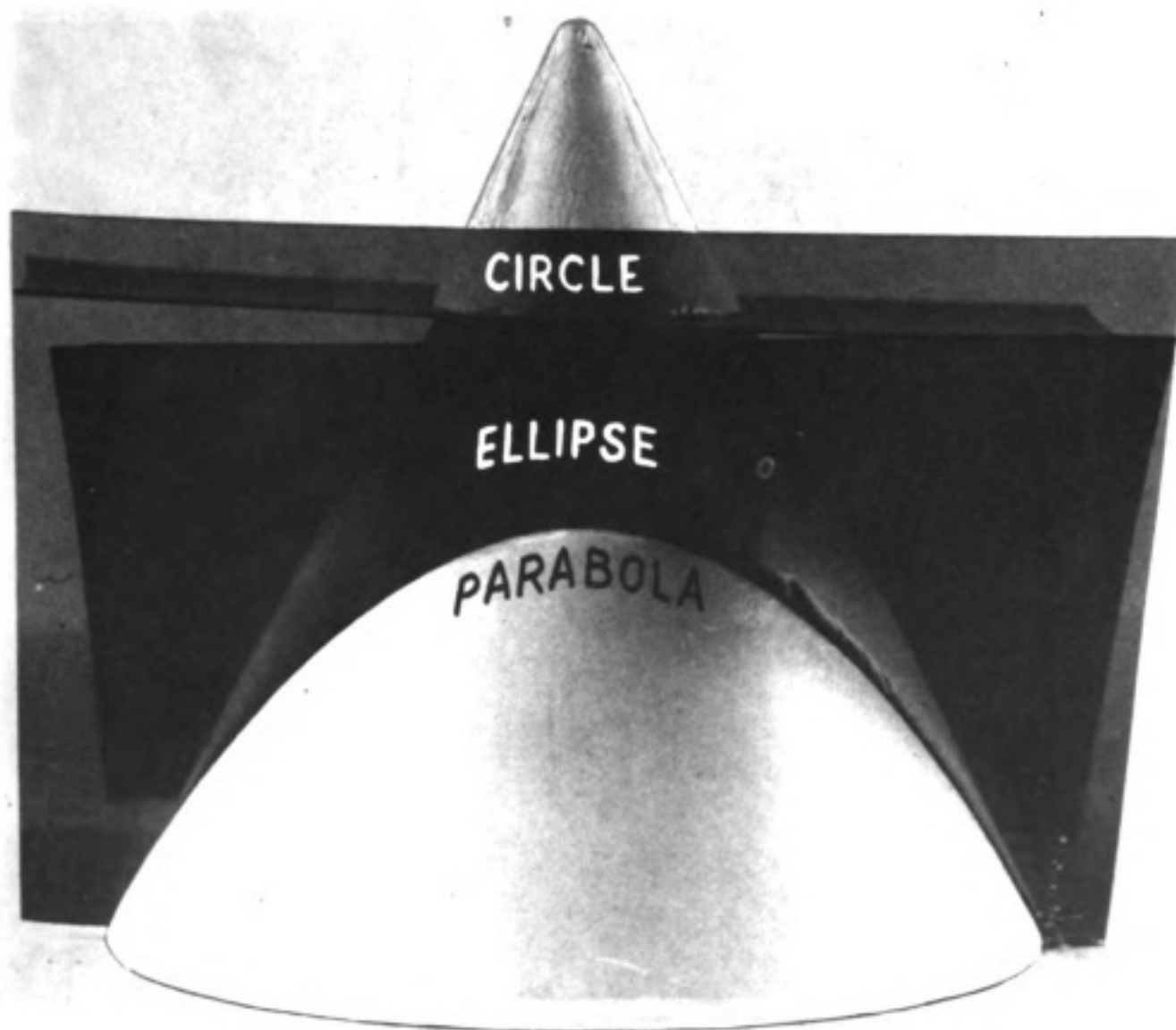


FIGURE 3

**SAME MODEL AS IN FIG. 1 AND FIG. 2, BUT FROM DIFFERENT CAMERA
ANGLE.**

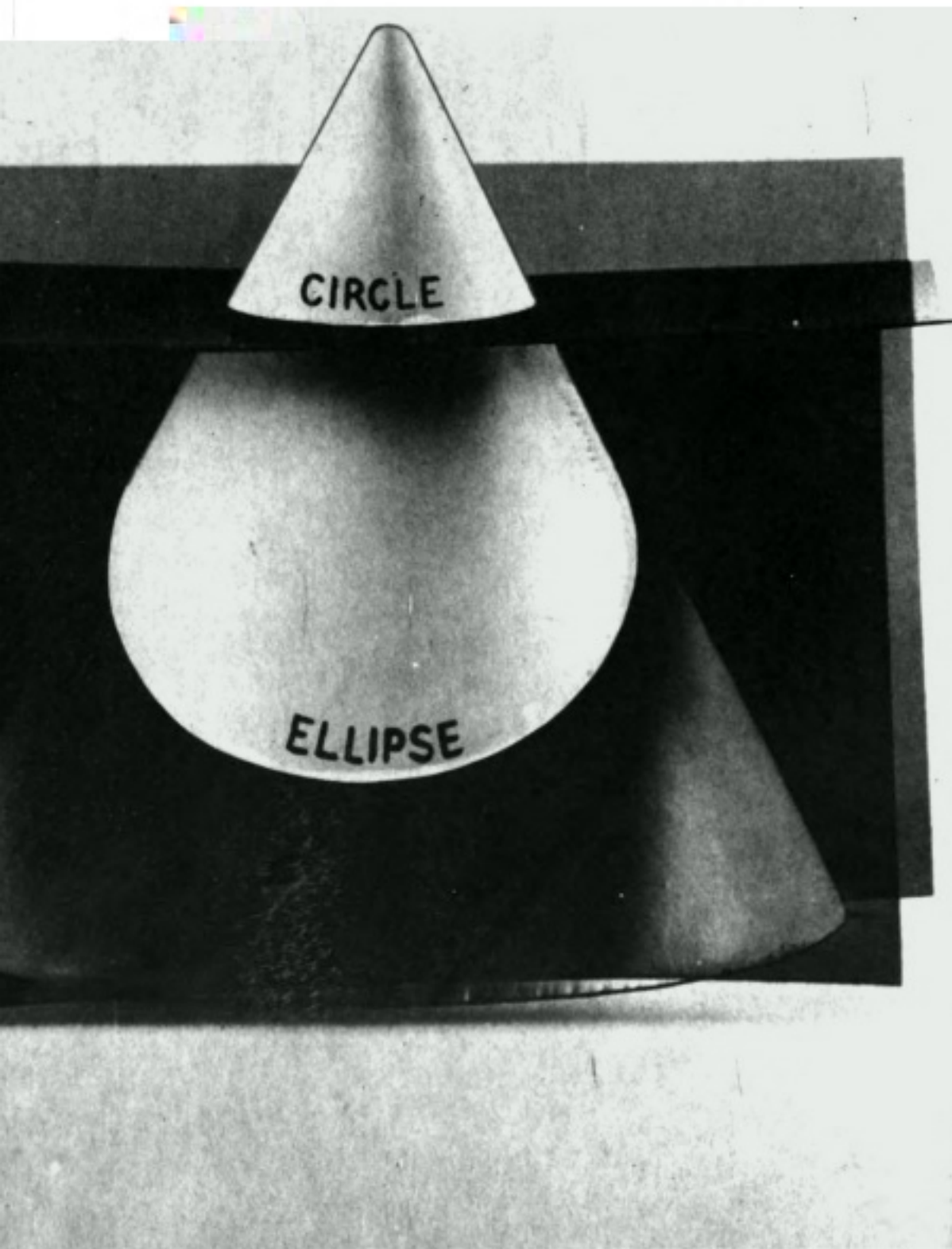


FIGURE 4

**3 FIG. 1, FIG. 2 AND FIG. 3, BUT FROM DIFFERENT CAMERA
OF CIRCLE AND ELLIPSE ARE QUITE EVIDENT.**



**SAME MODEL A
ANGLE, PARTS**

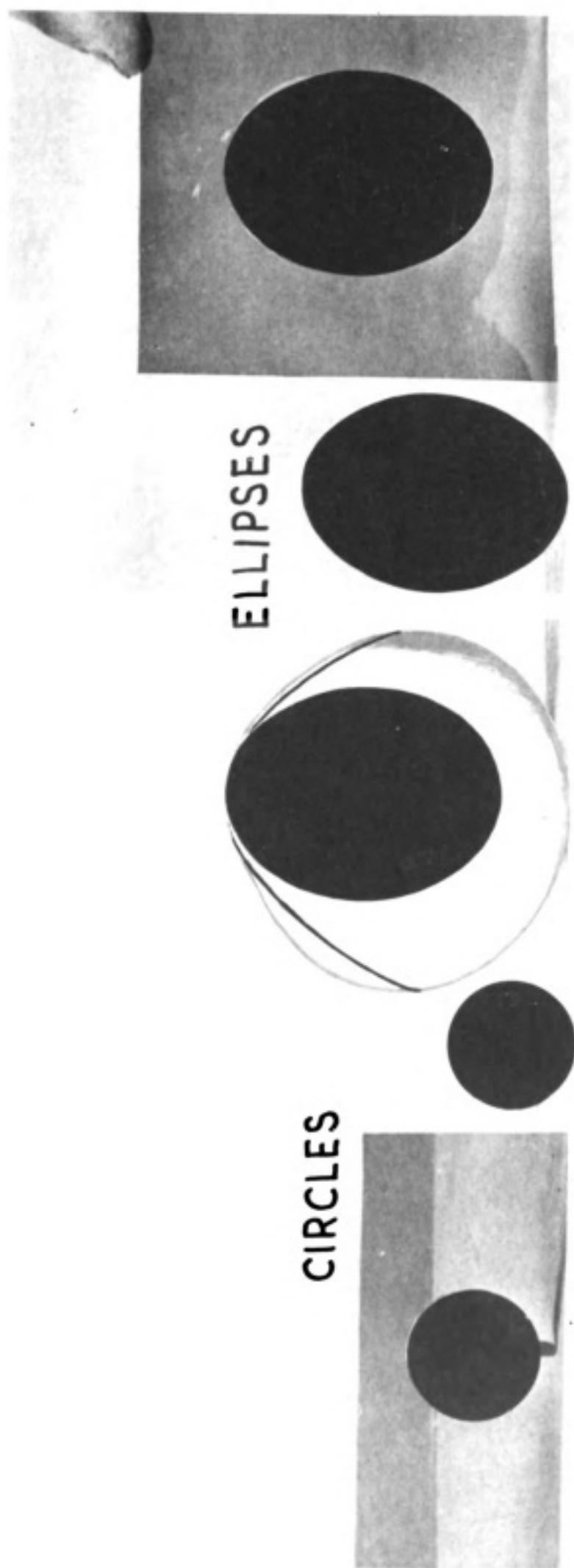


FIGURE 5

MODEL FIG. 1 IN DISSECTION. AT FAR LEFT IS INTERSECTING PLANE ITSELF SHOWING CIRCULAR AREA. THEN NEXT TO IT ON THE RIGHT IS BASE OF CIRCULAR TOP SECTION OF MODEL. IN THE CENTER IS THE ELLIPTICAL SECTION OF MODEL. AT THE RIGHT ARE THE ELLIPTICAL AREA ON INTERSECTING PLANE AND ELLIPTICAL AREA OF THE MODEL PIECE TAKEN OFF FROM THE MODEL AT CENTER.



FIGURE 6

THE DISSECTED MODEL OF FIG. 1, SHOWING PARABOLA ON INTERSECTING PLANE, ON THE MAIN PORTION OF THE MODEL AND REMOVED MODEL PIECE.

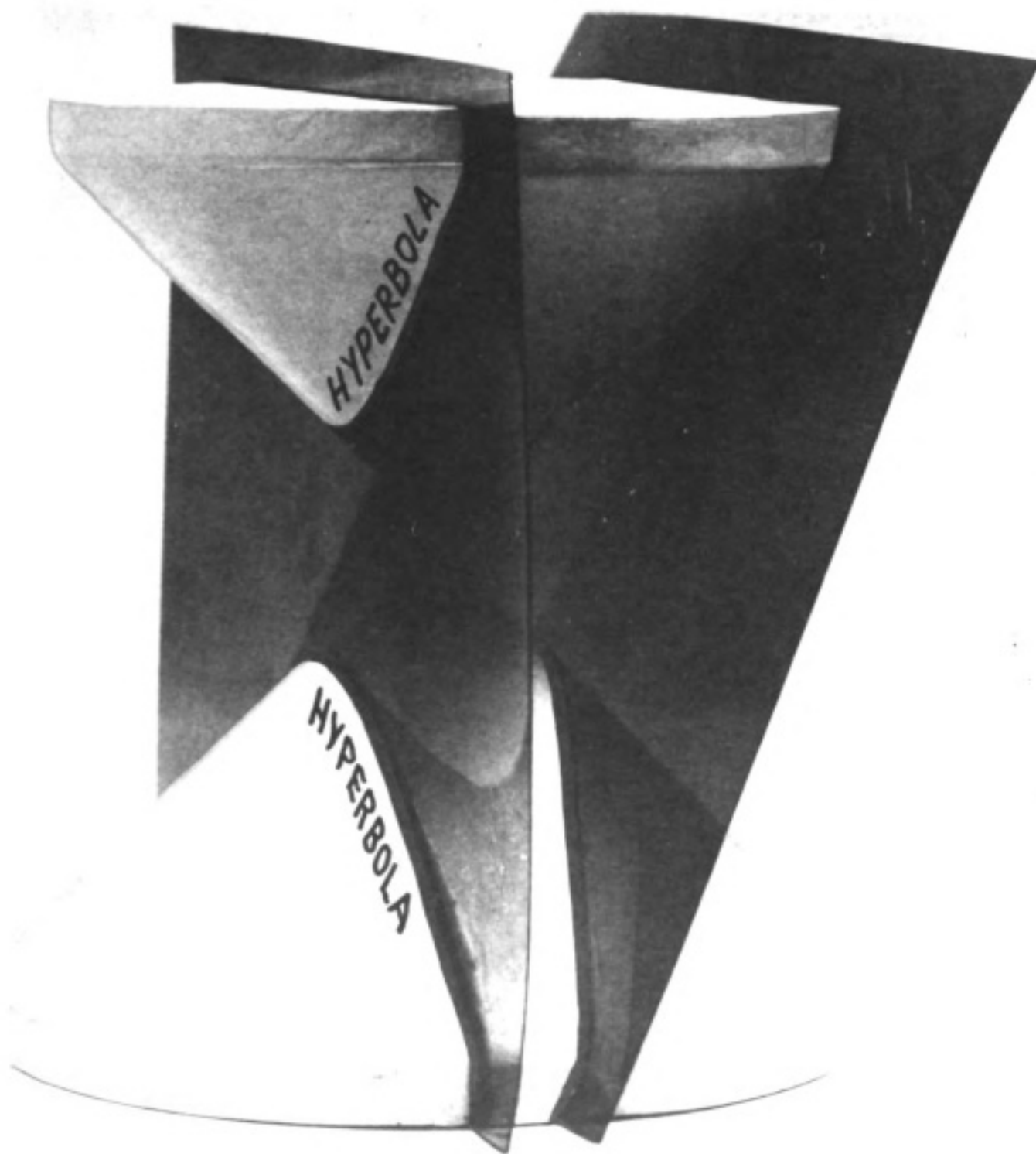


FIGURE 7

MODEL AND INTERSECTING PLANES GIVING HYPERBOLAS IN TWO NAPPES. THE COMMON APEX ANGLE IS 90° . THE BASE ANGLES OF BOTH NAPPES ARE 45° . THE LEFT INTERSECTING PLANE IS NOT PARALLEL TO COMMON VERTICAL AXIS. THE HYPERBOLAS ON THE RIGHT SIDE ARE NOT READILY EVIDENT.

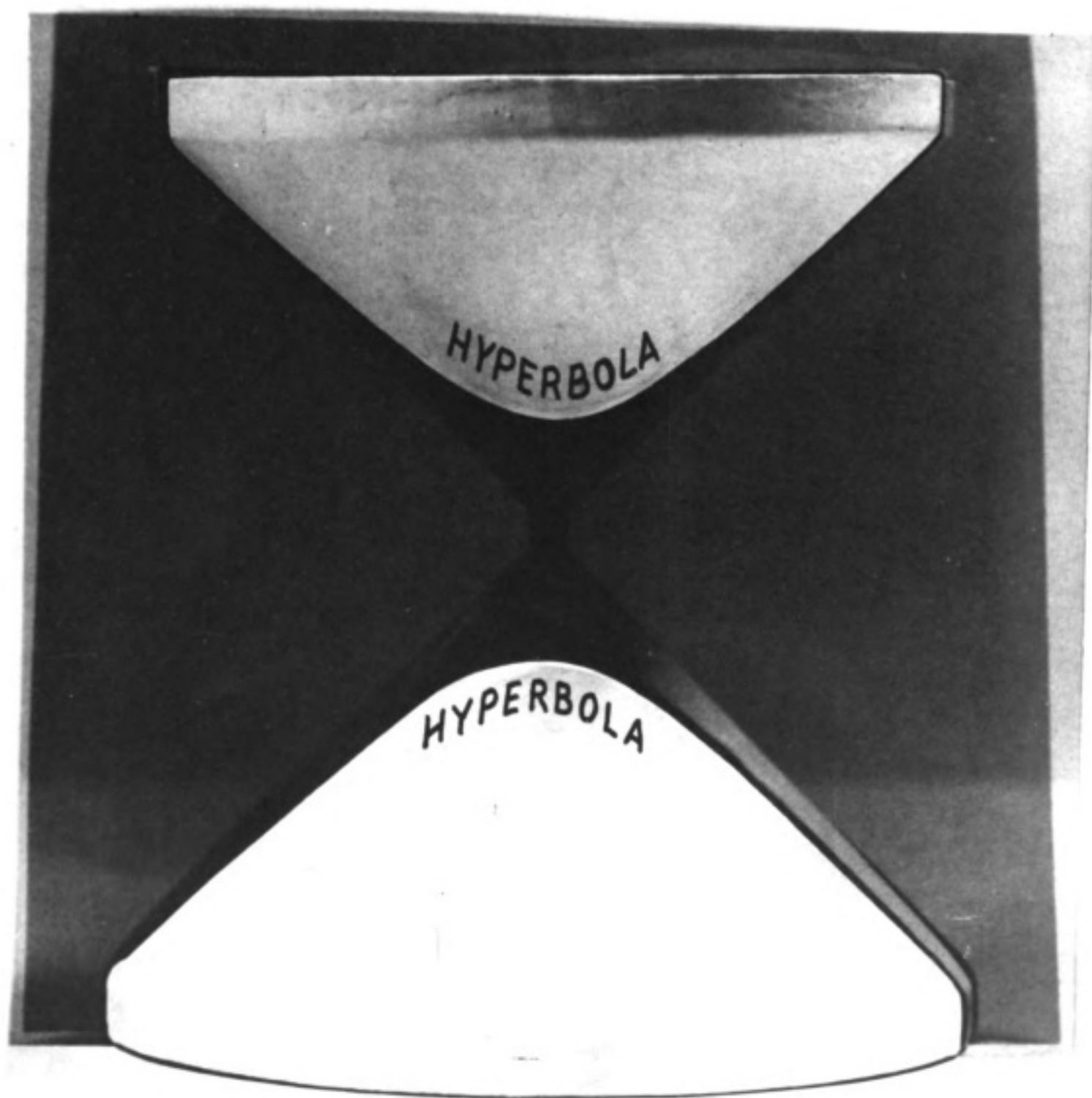


FIGURE 8

EQUILATERAL HYPERBOLAS OBTAINED FROM FIG. 7 BY PARALLEL INTERSECTING PLANE. ECCENTRICITY OF BOTH EQUILATERAL HYPERBOLAS EQUALS 1.41.



FIGURE 9

EQUILATERAL HYPERBOLAS OBTAINED BY DISSECTION MODEL AS SHOWN IN FIG. 8. AT LEFT ARE CURVES ON INTERSECTING PLANES; IN CENTER ARE CURVES OF MAIN SECTION OF MODEL; AT RIGHT ARE THE CURVES OF REMOVED PIECES OF MODEL.

THE UNIVERSITY OF CHICAGO
LIBRARY
1100 EAST 58TH STREET
CHICAGO, ILL. 60637
TEL. 773-936-5000
FAX 773-936-5001
WWW.CHICAGO.EDU



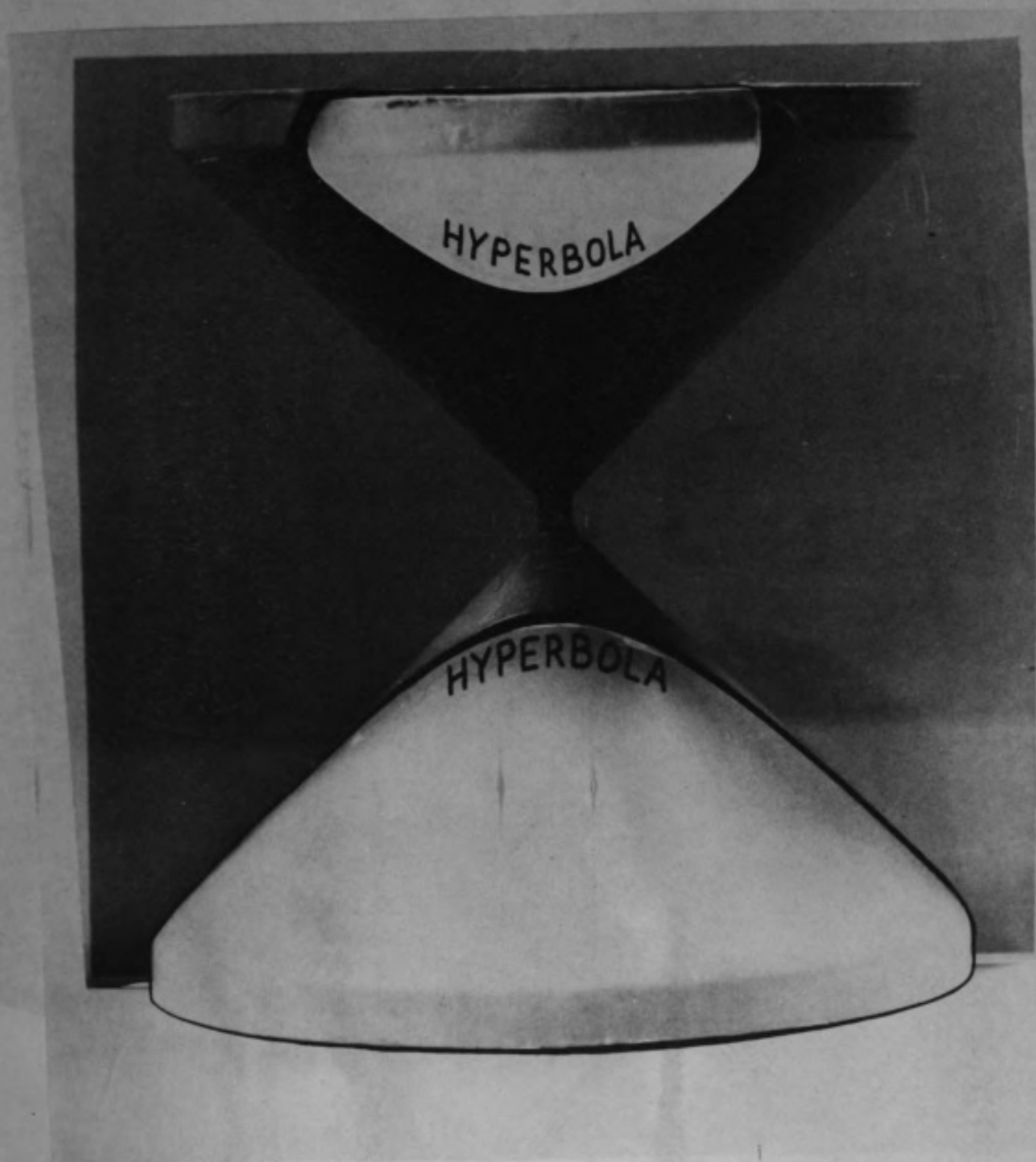


FIGURE 10

HYPERBOLAS OBTAINED FROM FIG. 7 WHEN INTERSECTING PLANE IS OTHER THAN PARALLEL TO COMMON VERTICAL AXIS. THEIR ECCENTRICITY IS EQUAL, AND LESS THAN 1.41 AND GREATER THAN 1. THE CURVES ARE THE SAME, BUT ARE OUTLINED TO DIFFERENT LENGTHS ALONG THEIR PATH.

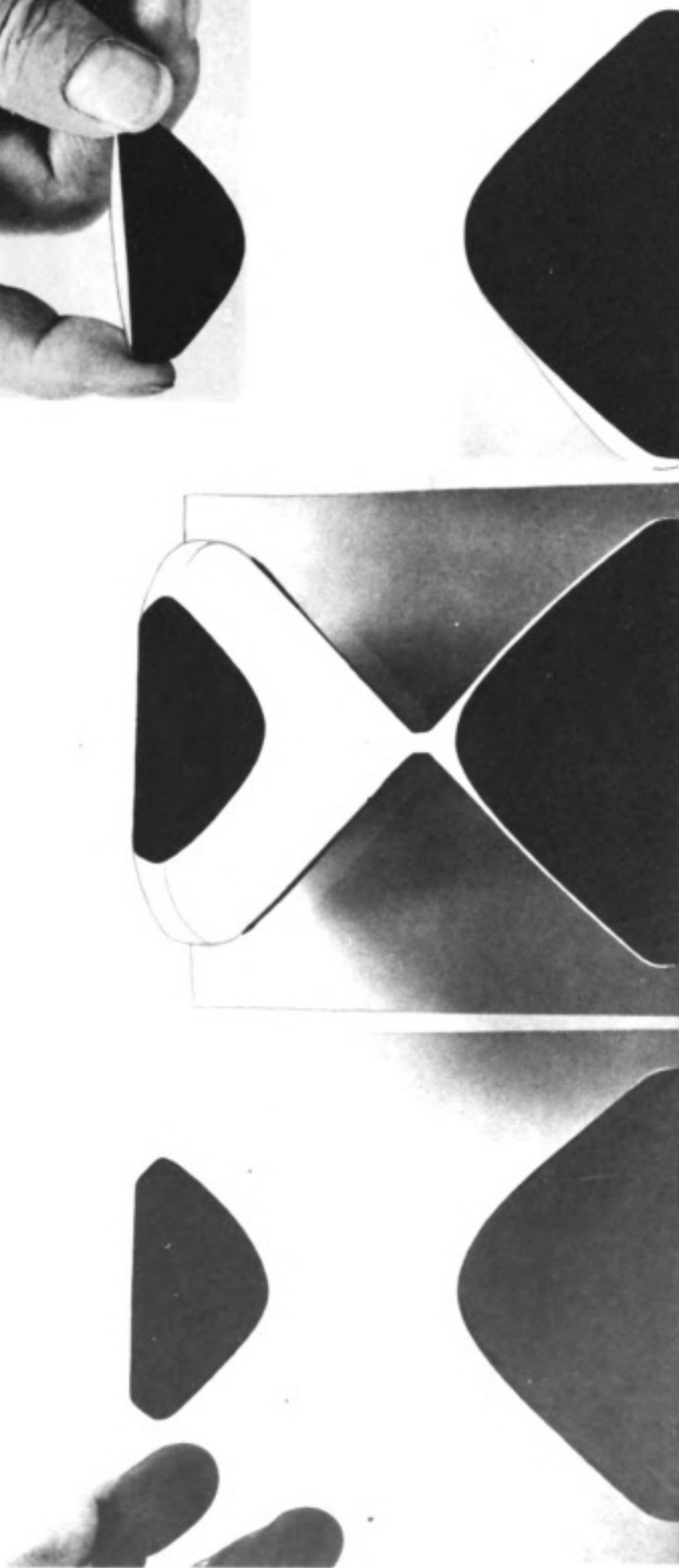


FIGURE 11

HYPERBOLAS OF DIFFERENT LENGTH BUT SAME ECCENTRICITY AS OBTAINED FROM FIG. 10. AT LEFT ARE THE CURVES AS OUTLINED ON NON-PARALLEL INTERSECTING PLANE; AT CENTER ARE THE CURVES OF MAIN PORTION OF MODEL; AT RIGHT ARE THE CURVES AS PROVIDED BY THE REMOVED SECTIONS OF THE MODEL.

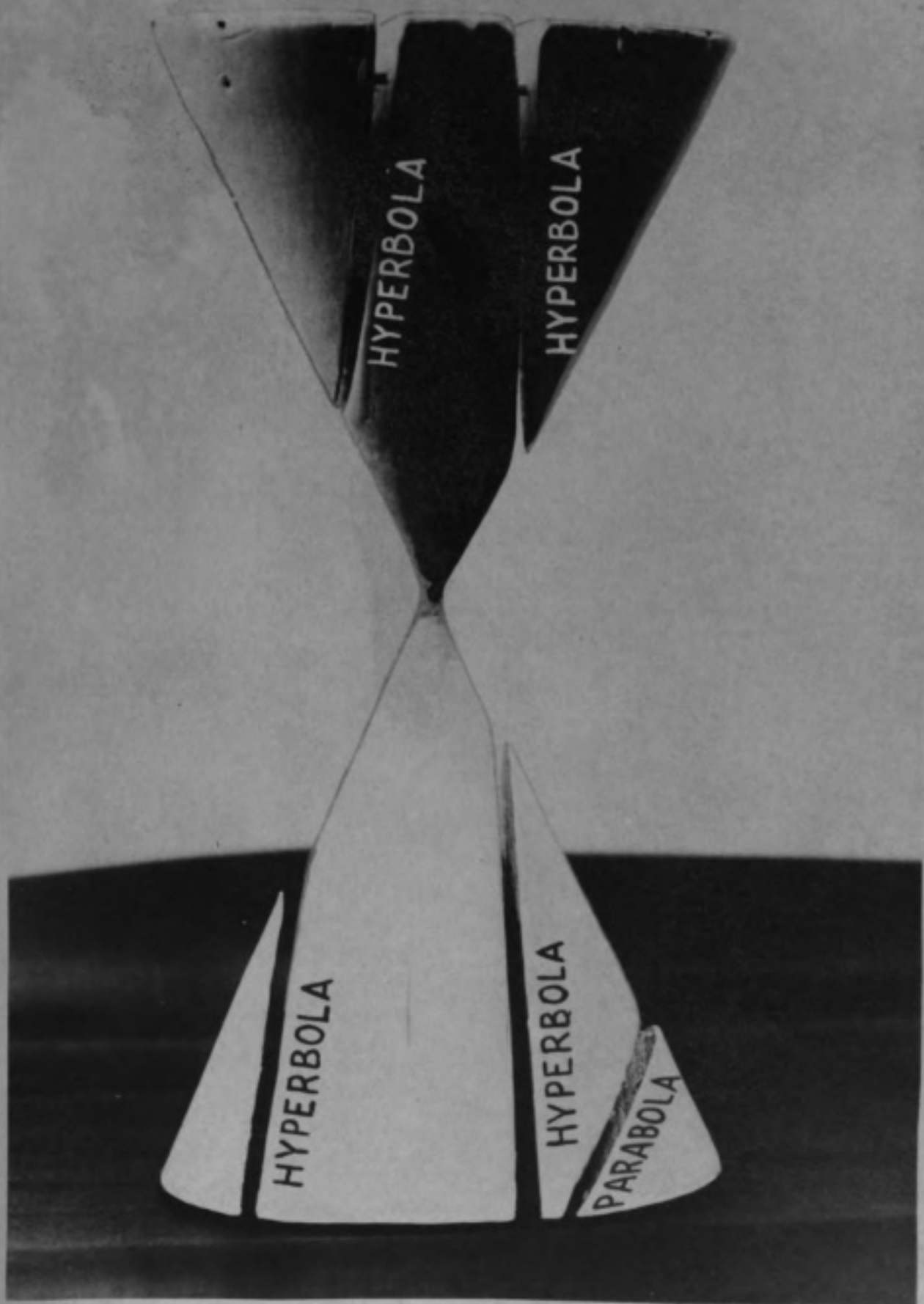


FIGURE 12

COMMON APEX ANGLE IS LESS THAN 90 DEGREES. A SMALL PARABOLA IS AT LOWER RIGHT-HAND CORNER. COMMON INTERSECTION PLANE AT RIGHT IS PARALLEL TO COMMON VERTICAL AXIS. THE HYPERBOLAS THUS CREATED HAVE EQUAL ECCENTRICITY AND LESS THAN 1.41. INTERSECTING PLANE AT LEFT IS ASLANT WITH COMMON VERTICAL AXIS PROVIDING TWO HYPERBOLAS WITH EQUAL ECCENTRICITY WHICH IS LESS THAN THAT OF OTHER HYPERBOLAS.

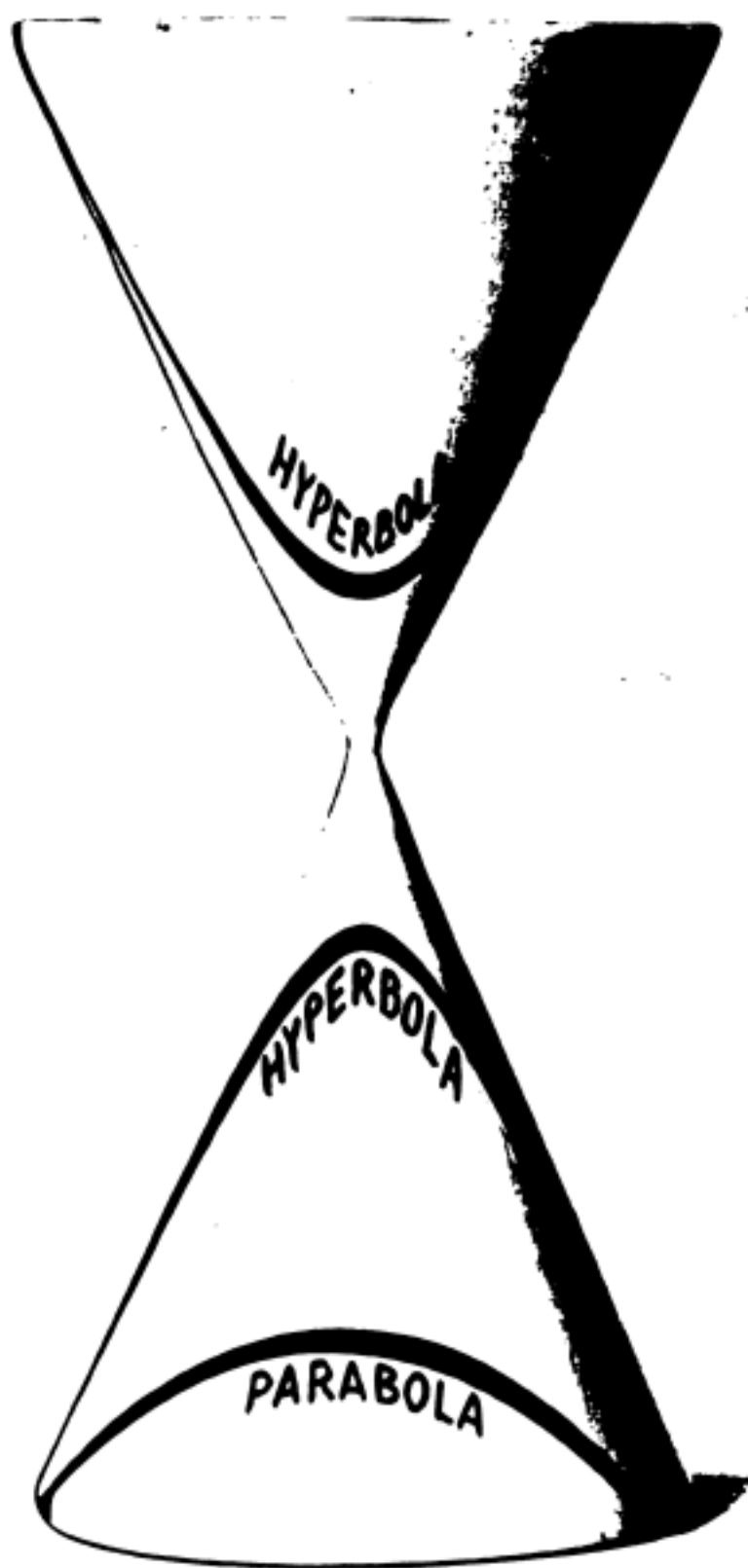


FIGURE 13

SAME AS FIG. 12 FROM DIFFERENT ANGLE. AT THE BOTTOM IS PARABOLA. THE TWO HYPERBOLAS SHOWN (CREATED BY PLANE PARALLEL TO PRINCIPAL AXIS) HAVE EQUAL ECCENTRICITY. SINCE COMMON APEX ANGLE IS SMALL, THE TWO HYPERBOLAS ARE ALMOST PARABOLAS.

CHAPTER 2

The Circle, Sine and Cosine, Areas

THE CIRCLE, AREAS, TRIGONOMETRY, SINES AND COSINES

The Greeks were splendid geometers and architects. During the thousand years before the Christian era, Euclid and others made many outstanding contributions to mathematics. The validity of their many theorems in geometry have stood the tests of time. It is interesting to observe they approached most of their geometric problems from the standpoint of areas. Problems connected with geometric figures bounded by straight lines and curved lines were always attacked from the point of view of areas. It was indeed an excellent doctrine which provided understanding and insight into the validity of their theorems.

RIGHT TRIANGLES

According to Pythagoras, born 582 B.C., the area of the square on the hypotenuse of a right triangle is equal to the sum of the areas of squares on the two sides of the triangle.

This simple but basic truth stirs the interest of the young school boy in the 4th and 5th grade. The boy has accepted arithmetic without a great deal of excitement. For him, arithmetic has been a matter of rote memory. But when his arithmetic gets tied up with a bit of geometry, he is likely to get one of his first intellectual thrills.

The boy has known that he saved time and steps when he cut "cater-cornered" across an empty lot. But when, for the first time, he learns that the hypotenuse of a square is only 1.41 times as long as one of the sides of the square, he cuts "cater-corner" across the empty lot on the way to school with great zest. Instead of walking two blocks around, he walks only 1.41 blocks on the hypotenuse. He has "saved" .59 of a block. Maybe for the first time in his young life he finds it is fun and exciting to think.

THE PYTHAGOREAN THEOREM

The boy's new found knowledge has come from his teacher's fine demonstration of the Pythagorean theorem in a manner about as follows:

Particularly the Pythagorean theorem is exciting, if we make a drawing (as shown in Figure 14) and then take several looks at the drawing. One does not say that the square of the hypotenuse is equal to the sum of the squares of the two sides but rather say, the area of the square made on (emphasis on the "on") the hypotenuse is equal to the sum of the areas of the squares made "on" the other two sides.

As seen in drawing, there is a right angled triangle ABC

$$AB = 3 \text{ inches}$$

$$BC = 4 \text{ inches}$$

$$\text{Area} = 5^2 = 3^2 + 4^2 = 25 \text{ square inches}$$

The area of the square "on" the hypotenuse AC is $ACDE$ and is equal to 25 square inches in area.

The area of the square "on" the vertical leg BC is $BCFG$ and is equal to 16 square inches in area.

The area of the square on the horizontal leg AB is $ABIH$ and is equal to 9 square inches.

Now if we invoke some simple algebra, we get some more simple and basic ideas about the truth and reality of Pythagoras' theorem.

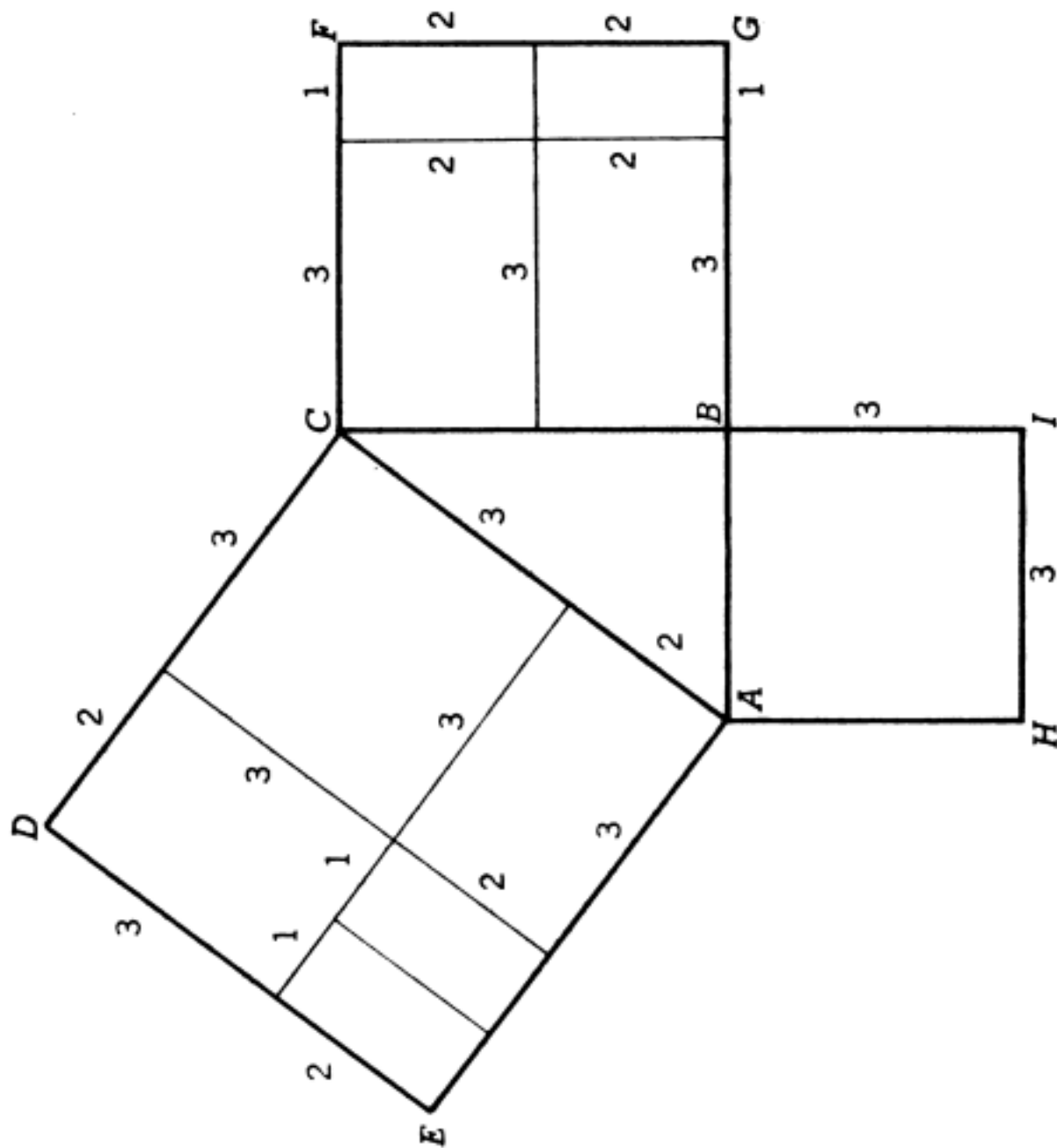
Let us say that $5 = (3 + 2)$.

By simple algebra $(a + b)^2 = a^2 + 2ab + b^2$

$$5^2 = 3^2 + 2 \cdot 3 \cdot 2 + 2^2 = 25$$

$$5^2 = 9 + 12 + 4 = 25$$

* This expression $2 \cdot 3 \cdot 2$ means 2 times 3 times 2. The dot in mid-position is used as a notation to indicate multiplication. The notation is frequently used in later pages. It obviates use of the notation, $\times$.



Area of Square on AC = Sum of Areas of Squares on BC and AB

$$AC^2 = BC^2 + AB^2$$

THE PYTHAGOREAN THEOREM
FIGURE 14

The 5th grade teacher does not use the algebraic analysis for the boy, to be sure. The algebraic approach is part of the teacher's "know how" which enables her to do a good graphical job as per the following procedure.

Put the square on AB (3×3) over into the square on AC (5×5). This takes care of the 3^2 or the 9 in the above algebraic expression.

Now divide the square on BC (4×4) in 4 separate areas. Two of these latter areas are each 2 by 3, and the other two are each 2 by 1. This accounts for the $2 \cdot 3 \cdot 2$ and the 4 respectively in the above. ($12 + 4 = 16$.) It will be seen by inspection how the separate areas of the square on BC fit in to the remainder of the square on AC after the square on AB (3×3) was put into the square on AC . Altogether $5^2 = 3^2 + 4^2$ has been shown as a valid fact by graphical means.

Pieces of cardboard can be cut properly to do this job. It is a most effective scheme to have six pieces of cardboard and use them in a sort of jig saw puzzle game fashion.

So the above sketch, simple algebra, a jig saw puzzle and simple analysis in general, we grown ups too readily see that Pythagoras was right. One has a convincing proof of the ancient theorem. The pieces of cardboard constitute a training aid.

The boy's one block square vacant lot experience now begins to "make sense." The diagonal is $\sqrt{1^2 + 1^2} = \sqrt{2} = 1.41$. He learns why he saves .59 of a block.

CIRCLES

Consider the circle shown in Figure 15, page 24.

By simple definition, a circle is a curve, every point of which is equidistant from the center O . This equal distance is R , the radius of the circle.

PYTHAGOREAN THEOREM APPLIES TO ALL POINTS ON CIRCLE

Consider the point A . Its distance from the center is R . From A , straight down to horizontal reference line X_1X , the distance is AB . From B to center of circle, the distance is BO .

CIRCUMFERENCE AND AREA OF CIRCLE

The circumference of a circle and the calculation of its area offers another exciting episode in the life of the grown up boy probably now in eighth grade. He learns that the circumference of a circle is 3.1416 times as long as the diameter. $C = \pi D = 2\pi R$. He meets π (pi) for the first time. He is intrigued by 3.14159 as the value of π . He winds a string around a cylinder and finds the string is about $3\frac{1}{7}$ times as long (3.14) as the diameter. He gets a kick out of his new found knowledge.

Emulating the Greeks, his curiosity immediately prompts him to ask about the area of a circle. He likely has learned that the area of an isosceles triangle is the base times $\frac{1}{2}$ the height. The circle can be thought of as many very slender isosceles triangles. The height of all of them is half of the radius. The sum of the bases of all the triangles is the circumference or π times the diameter (πD); π times $2R = 2\pi R$.

$$\text{Area of circle} = 2\pi R \text{ times } \frac{R}{2} = \pi R^2$$

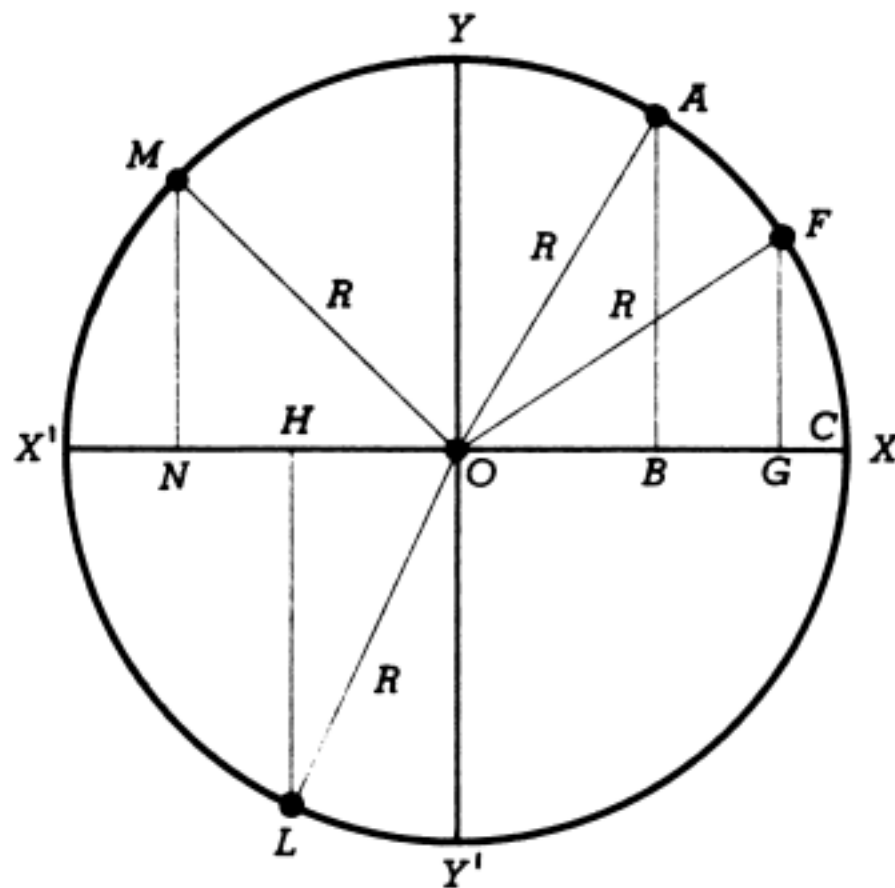
This again gives the boy a thrill. The boy does not realize that he has been introduced to the beginnings of integral calculus. Later he will get another thrill when he studies calculus and finds out how areas bounded by curves (other than circles) and straight lines also can be readily determined. In every case, one needs to know the equation of the curve at hand. He also will be interested to learn about an instrument called the planimeter which measures area, whether the area be regular or irregular in shape.

CIRCLES AND PYTHAGORAS

Triangle AOB is a right triangle in the accompanying drawing figure 15 and the square on AO is equal to the sum of the squares on AB and OB , a la Pythagoras. So $AO^2 = AB^2 + OB^2 = R^2$.

Now take another point beside A on circle, namely point F . In this case $OF^2 = FG^2 + OG^2 = R^2$. Now take

CIRCULAR SINES AND COSINES



$$\overline{FG}^2 + \overline{GO}^2 = \overline{AB}^2 + \overline{BO}^2 = \overline{MN}^2 + \overline{NO}^2 = \overline{HL}^2 + \overline{HO}^2 = Y^2 + X^2 = R^2$$

$$\text{SINES} = \frac{\text{ORDINATES}}{\text{HYPOTENUSE}}$$

$$\frac{\text{CONSTANT}}{\text{HYPOTENUSE}} = R$$

$$\text{COSINES} = \frac{\text{ABSCISSAE}}{\text{HYPOTENUSE}}$$

$$\left(\frac{Y}{R}\right)^2 + \left(\frac{X}{R}\right)^2 = 1$$

$$\text{Sin}^2 + \text{Cos}^2 = 1$$

PYTHAGOREAN THEOREM APPLIES TO ALL POINTS ON CIRCLE

FIGURE 15

a third point on circle, namely M . In this case $OM^2 = MN^2 + ON^2 = R^2$. For every point the Pythagorean theorem applies.

The three points A , F and M are of course all at a distance of R (radius of circle) from O . Every point on the circle is likewise at a distance of R from O .

ORDINATES AND ABCISSAS

In geometry and in so called analytical geometry, the straight line distance from any point on the circle to the horizontal reference line or X axis is given the name "ordinate." The ordinate is always the length of line parallel to vertical reference line or Y axis. Furthermore, this ordinate distance is conventionally designed by Y ; it is distance along the Y axis. Ordinates are Y 's.

The distances from the center of the circle (O , the intersection of the X and Y axes) to B , G or N on X axis are called X 's or abscissas. The X distance and the Y distance are called the coordinates of the particular point on the circle.

EQUATION OF A CIRCLE

So we can put all the ideas and words that have been used in the above paragraphs into the symbols of a simple equation

$$X^2 + Y^2 = R^2$$

$$(\text{Abscissa})^2 + (\text{Ordinate})^2 = (\text{Radius})^2$$

This $X^2 + Y^2 = R^2$ is called the equation of a circle. It symbolically expresses simple, plausible and readily understandable truths and facts about every point in the circumference of a circle. Emphasis is to be made that the equation states a simple fact for each and every point in the circumference. One would say the equation of a circle has been simply derived by just plain, ordinary common sense plus, of course, the use of the Pythagorean theorem.

TRIGONOMETRY

Trigonometry is a branch of mathematics having to do primarily with triangles. All triangles have tri-angles, namely three angles. Gonia is a Greek word meaning angle or corner. The word gonia also involves the idea of a wedge or a cleavage or a cutting. Metron is also a Greek word meaning measure.

An angle is formed by two intersecting lines forming a corner or a wedge. Add all these ideas and interrelations thereof and one gets trigonometry, a branch of mathematics having to do with the measurement and interrelations of angles, corners or wedges as they appear in triangles.

If now we take the equation of a circle

$$Y^2 + X^2 \text{ or } X^2 + Y^2 = R^2$$

and work a little algebra on it, namely divide both sides of the equation by R^2 one comes up with the following form of the equation for a circle.

$$\frac{Y^2}{R^2} + \frac{X^2}{R^2} = 1.$$

or

$$\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1.$$

$\frac{Y}{R}$ is of course the ordinate AB for point A in above sketch divided by the radius, R . $\frac{X}{R}$ is the abscissa OB for point A divided by the radius.

SINES AND COSINES

At this point the "trig" man says, "I will call the ratio of the ordinate (AB or Y) to the radius, (R) the sine of the angle (AOB) formed at the center of the circle by the lines AO and BO ." For point M on the circle, the ratio of the ordinate (MN or Y) to the radius (R) will be called likewise the sine of the angle (MON) formed at O by the lines MO and NO . There is another angle MOX herein involved formed by lines MO and OX , but the sine of angle MON is equal to the sine of MOX . Angles MON and MOX have the same ratio of ordinate (Y) to abscissa (X). In general, the sine is the ratio of an ordinate to the radius R , namely $\frac{Y}{R}$.

Now as regards the abscissa for point A , the ratio of the abscissa distance OB (X) to the radius R is called by the "trig" man, the cosine of the angle AOB . For the angle MON and the angle MOX the cosine is NO divided by MO . In general, the cosine is the ratio of an abscissa to the radius R , namely $\frac{X}{R}$.

So we can write the equation of a circle a third way, naming the angle involved in every case θ ; theta is a Greek letter. Theta is written θ .

$$\sin^2 \theta + \cos^2 \theta = 1.$$

or

$$\cos^2 \theta + \sin^2 \theta = 1.$$

ROTATION OF POINT IN CIRCUMFERENCE

Consider a point in the circle, as for example point C , and assume it goes around and around the circumference in a counterclockwise direction. The ratio of the ordinate of the point (as it swings round and round) to the radius (even when the point is below the X reference axis) is the sine of the angle.

If one would plot along the Y axis the numerical values of this sine or ratio $\left(\frac{Y}{R}\right)$ for all positions of the point, as it starts at point C (Fig. 15) and rotates counterclockwise, versus the angle of rotation plotted along the X axis, the sine curve (page 27) would result, Fig. 16.

SINE CURVE—A COMBINATION OF TWO SIMULTANEOUS MOTIONS

When a point moves in a circular path with uniform velocity, it will have a projected vertical motion up and down along the Y axis. This Y axis motion will be the perpendicular or orthogonal projection of the circular motion. The projected distances per unit of time (angle) will vary as the sine. As the point goes round and round in the circle, the Y axis motion will be "simple harmonic motion." The prongs of a tuning fork swing back and forth in simple harmonic fashion.

If this latter oscillatory motion is combined simultaneously with some constant velocity to the right along the X axis, the resultant path or curve will be a sine curve or sine wave. Such a sine curve is shown in accompanying Figure 16.

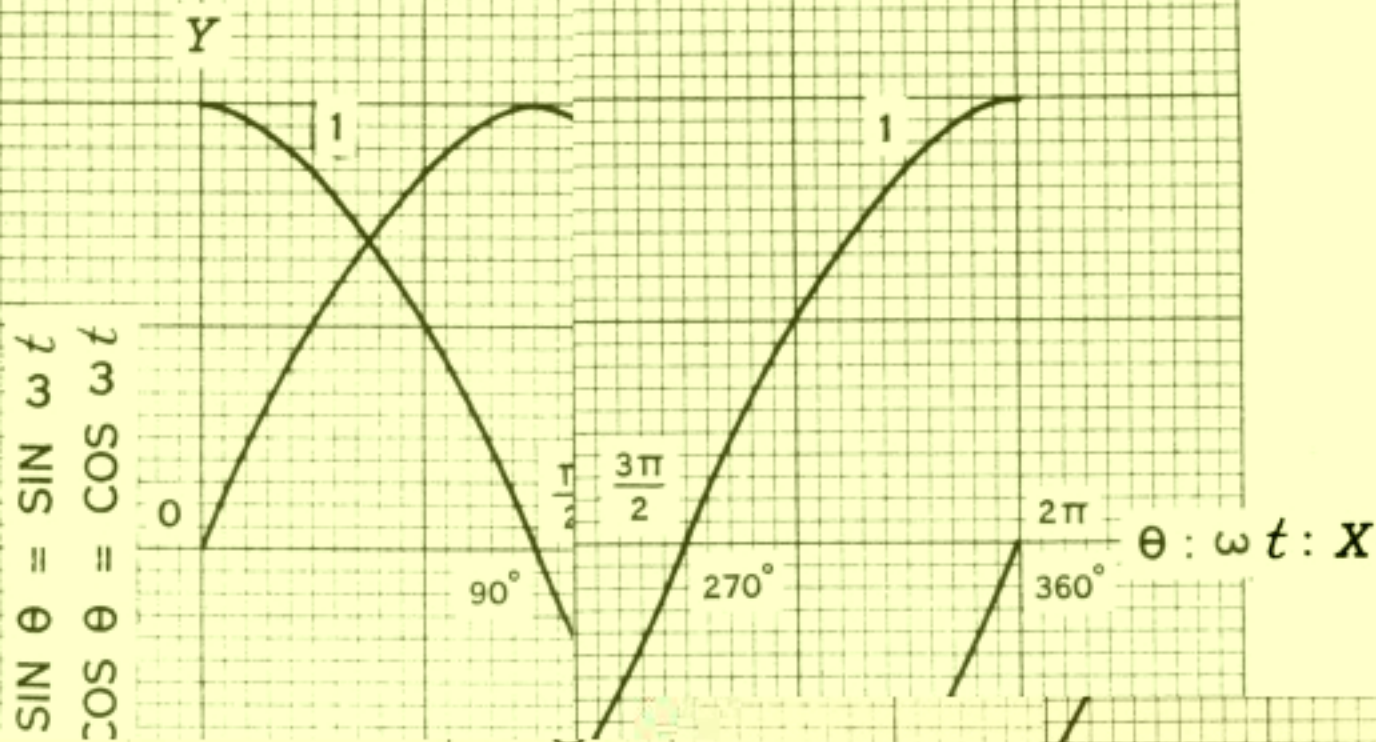
SINE CURVE AREA

It is of no particular importance but it is a point of interest and understanding to consider the area under one "loop" of a sine curve (cosine too). To examine the sine curve from the area point of view would be in keeping with the helpful Greek doctrine of thus considering matters relating to curves in general.

The area of one "loop" of a sine curve would have 3.1416 (π) for a base. If the maximum ordinate of the sine curve were called Y , then the area at issue is numerically equal to $2Y$. This means that with a base of 3.1416 units and a height of curve of unity (1), the area under curve would be 2 square units.

If a rectangle were built, as it were, about one loop of the sine curve with base of 3.1416 units and 1 unit high, the area of the rectangle would be 3.1416 square units. This means that the area under a single loop of the sine curve is about $\frac{2}{3}$ the area of the "corresponding" rectangle.

SINE CURVE RATE OF CHANGE (s t



t = TIME
 f = FREQUENCIES, CYCLES
 T = PERIOD = $\frac{1}{f}$

0 COS $2\pi ft$
 0 SIN $2\pi ft$

0 COS $\omega t = 2$
 0 SIN $\omega t = 2$

A rectangle is indicated by the dotted lines Fig. 16. The height of the maximum ordinate in the sine curve of drawing is 2 units. The base is really 3.1416 units. The area under curve is 4 square units. However for convenience and ease in plotting the 3.1416 is arbitrarily represented by a scale of plotting by which 3.1416 is 3.000. Adoption of a scale is entirely a matter of one's choice. So 3 actually represents 3.1416 in the plotting.

Even with the arbitrary scale adopted, the area of the rectangle on figure 16 is 6 square units (instead of 6.2832). The point is nicely illustrated that the area under the sine curve is 4 square units and is $\frac{2}{3}$ of the area of the corresponding rectangle. By inspection of the curve area on figure 16, one can see that the area could add up to 4 square units. Of course the 6 square units of rectangular area are evident.

The interesting fact that the area under one loop of a sine curve, and a cosine curve too, is numerically equal to twice the maximum ordinate ($2Y$) is provided by a splendid "assist" by that branch of mathematics called integral or "adding up" calculus.

COSINE CURVE

If one would plot the ratio of abscissa to the radius for every position in the circle as the point rotates counter-clockwise from beginning position C and around back to C , Figure 15, as related to uniform linear velocity at right angles one would get a cosine curve.

The accompanying cosine curve shows that the point has rotated in circle from position C counterwise for 2 revolutions.

It is to be remembered and emphasized that, whatever curve (sine or cosine) one draws, the $\cos^2 \theta + \sin^2 \theta$ always equal 1; $\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1$.

In the realm of that branch of mathematics called differential calculus, the sine and cosine curves of Figure 16 illustrate basic and simple truths. The differential calculus says that the derivative, the time rate of change $\left(\frac{dy}{dx}, \frac{d\theta}{dt}, \frac{dwt}{dt}\right)$ of the sine is the cosine. Likewise, the derivative, the time rate of change $\left(\frac{dy}{dx}, \frac{d\theta}{dt}, \frac{dwt}{dt}\right)$ of the cosine is the sine. Geometrically, the $\frac{dy}{dx}$ at all points along the curves is the slope or steepness of that curve at each point.

TWO SIMULTANEOUS MOTIONS

The sine curve (and the cosine curve) can very appropriately be thought of as the recurrent and repeated path of an object having two simultaneous motions. The curve is the resultant or combination of these two simultaneous motions. See accompanying drawing of sine and cosine, Fig. 16.

The left to right motion on X axis is a uniform velocity while the up and down motion on Y axis is called simple harmonic motion or sinusoidal motion. The uniform velocity along X axis is actually the uniform angular velocity (ω) (the Greek letter omega) of the point in its spin around the circle. The up and down motion on Y axis is simple harmonic motion which is a very common form of back and forth motion. If a weight is suspended from a spring and given a little "nudge" to oscillate up and down, the motion is S.H.M. A pendulum on a clock swings back and forth with S.H.M.

S.H.M. is readily seen to be the orthogonal or perpendicular projection on either X or Y axis of an object moving in a circle with constant angular spin or constant angular velocity (ω).

REPETITIVE NATURE OF CIRCULAR SINE AND CIRCULAR COSINE CURVE

A major and a most important point about circular sines and circular cosines is, that as the point goes round and round the resulting sine and cosine curves (as illustrated) show not only that the sines and cosines vary in value from 0 (zero) to 1 unity and then again to 0 (zero), and from (1) unity to 0 (zero) and then again to 1 (unity) respectively but also that the process is repeated over and over. The process is cyclic, repetitive, recurrent, periodic.

This fact of periodicity or recurrence is not true of hyperbolic sines and cosines to be discussed later.

CIRCLE AND TWO SIMPLE HARMONIC MOTIONS

The circle can be very appropriately be thought of as the result of two S.H.M. motion at right angles to one another with equal amplitude but 90° difference in phase or relative position at starting point and at every succeeding point.

The vertical S.H.M. component on "Y" axis is sine curve motion while the horizontal S.H.M. component on "X" axis motion is portrayed by a cosine curve, 90° difference in phase or relative position ahead of the "Y" axis motion. In other words, at the instant of time when a point starts to swing around the circle from its starting point "C," the ordinate or Y position (the sine) is zero but the abscissa or X position (the cosine) is a maximum of 1 (unity).

TWO S.H.M. MOTIONS AT RIGHT ANGLES

Imagine a pendulum swing from one's right to one's left with S.H.M. it would be out at extreme right, then in center and finally out at extreme left in its single journey from right to left.

When at extreme right (ready to start back to center) its X distance or abscissa is a maximum. At this initial instant of time, assume one gives it a push so it will also swing away from the viewer. The pendulum will now not only swing from right to left but also away from and toward the observer. If the push is just the right amount, the pendulum "bob" will go in a circle. If its right to left swing distance and vice versa is the same as its "swing away" distance and its "swing toward" distance and vice versa, the pendulum "bob" will move in a circle.

Again it is to be stated that a circle can be thought of as two S.H.M.'s at right angles with a phase difference of 90° . The phase difference expressed in time is one fourth of the period of vibration; $T/4$.

When one started the second S.H.M. for the "bob", so it would move "away" and later "toward", we can say that time was zero (0). However, at $t = 0$, the X axis distance from middle to extreme right was already a maximum but the "away" distance was zero. The motion along X axis may be validly thought of therefore as varying as the cosine and is 90° ahead in phase of the Y axis motion which is varying as the sine. The Y axis motion (the away and toward motion) started just as we gave the bob a push, so it was, as far as Y axis is concerned, at zero (0) distance when $t = 0$. So again the Y axis is "a la" sine motion and the X axis motion is "a la" cosine motion. Both are a la simple harmonic motion. See Figure 17.

If an observer were a considerable distance out in front and not able to look up or down at the pendulum "bob" moving in a circle, the observer would think that the pendulum was merely moving right and left and vice versa in S.H.M. If another observer were way out to the right and not able to look up or down at the pendulum bob moving in a circle, the second observer would think that the pendulum was merely moving to his right and left and vice versa in S.H.M.

ELLIPTICAL PENDULUM

The accompanying chart Fig. 17, was made by two simple harmonic motions at right angles and unequal amplitude and 90° difference in phase. No. 1 curve portrays north and south motion of pendulum and is a sine curve. All points shown on ellipse were located by referring to a point on the sine curve and also to a point on the cosine curve. These points are numbered 1, 1, 1 and 2, 2, 2 etc.

No. 2 curve portrays east and west motion of pendulum and is a cosine curve and leads by 90° as seen at point S.

CIRCULAR PENDULUM

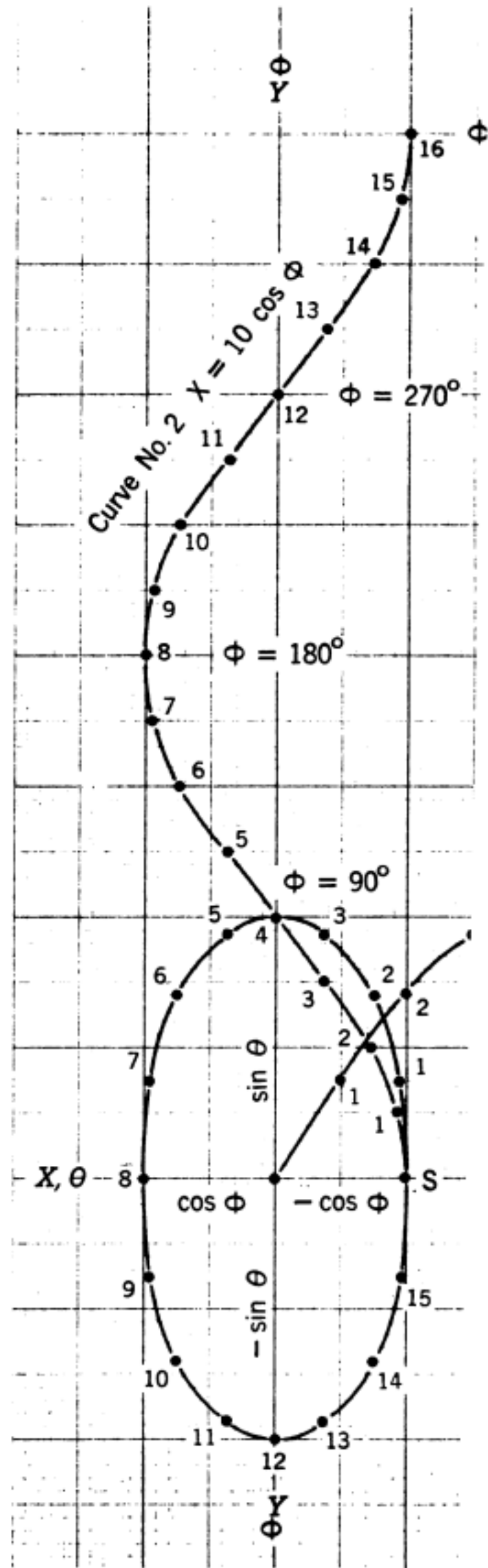
The accompanying chart Figure 18, was made by two simple harmonic motions at right angles and equal amplitude and 90° different in phase.

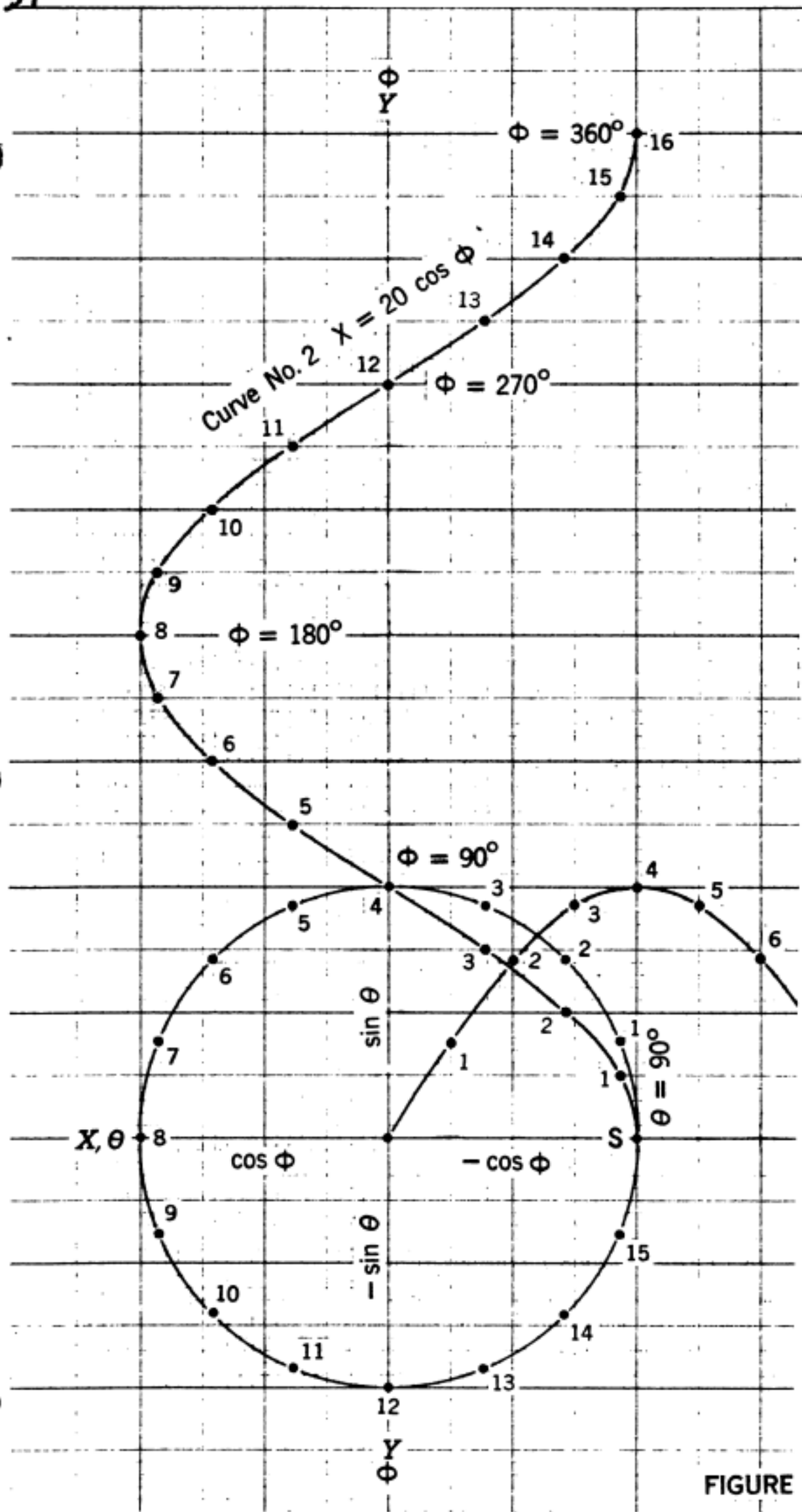
No. 1 curve portrays north and south motion of pendulum and is a sine curve.

No. 2 curve portrays east and west motion of pendulum and is a cosine curve and leads by 90° as seen at point S.

All points shown on circle were located by referring to a point on the sine curve and also to a point on the cosine curve. See correspondingly numbered points 1, 1, 1 and 2, 2, 2 etc.

The reader may wish to look up discussion of Blackburn's compound or double pendulum in a physics text book to get a further appreciation of two simple harmonic motions at right angles.





FIGURE

CHAPTER 3

The Ellipse

Later in this treatise the ellipse will be discussed somewhat in detail. But right now is a good time to say, that if the above mentioned push given at $t = 0$, were big enough to make the "away" distance swing and the "toward" distance swing on Y axis greater than the right and left swing distances on the X axis, the pendulum "bob" would describe an ellipse when looked down upon. The ellipse would have its longer "axis" along the Y coordinate axis and its shorter "axis" along the X coordinate axis.

However, one of the two observers in the same relative positions as before (way out in front) would think the "bob" was moving with S.H.M. along a straight line to his right and left. The other observer (way out to the right) would judge the "bob" was also merely swinging to his right and left along a straight line with S.H.M. The observer way out in front would report a swing of say 5 inches while the observer way out to the right would report a swing of say 7 inches.

THE CIRCLE; A SIMPLY DRAWN CURVE

After all, the circle is the easiest of all geometrical curves to draw. To draw even a straight line, one must have a carefully made ruler. But if one has a pin, a piece of string and a pencil, one can easily get a fine circle. All one has to do is to fasten the pin to one end of string and the pencil to the other end, then stick the pin in the paper and swing the pencil around with the stretched string guiding the lead pencil to make a fine circle.

As will be discussed in the section on the ellipse, it will be shown that an ellipse curve is almost as easily drawn by a very simple procedure. A parabola and a hyperbola are not so easily and simply drawn by simple tools like strings and pins. One has to use specially designed "tools" to draw a perfect parabola or hyperbola curve.

Finally again the equation of a circle can be written three ways.

$$X^2 + Y^2 = R^2$$

or

$$\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1$$

or

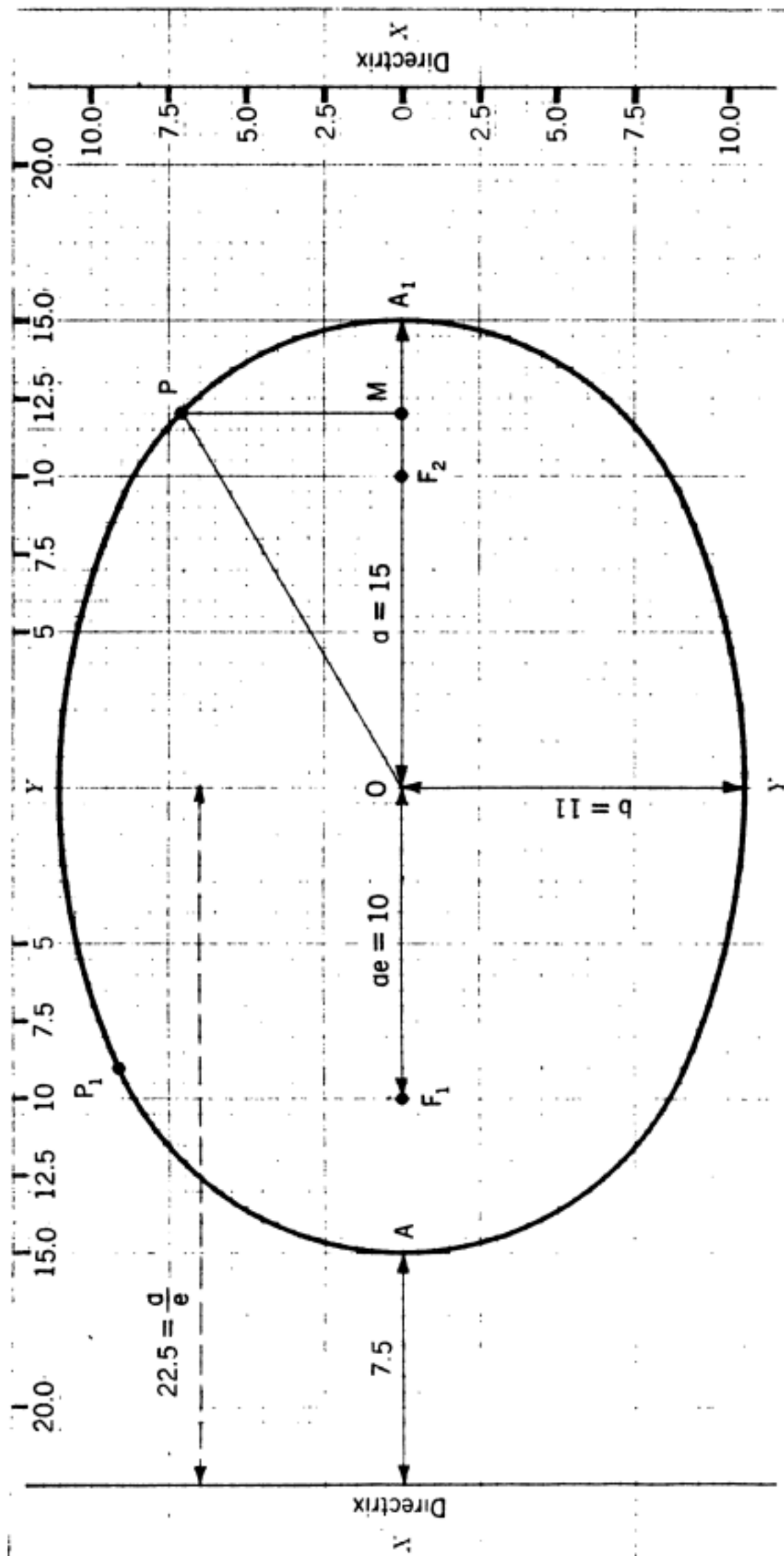
$$\cos^2 \theta + \sin^2 \theta = \sin^2 \theta + \cos^2 \theta = 1$$

THE ELLIPSE

The ellipse is another of the conic section curves and is a beautiful curve. It is a sort of lopsided or "squashed" circle. It has two axes, a short or minor one and a long or major one; a short and a long "diameter" as it were. It does not have a single central point or focus like a circle but has two central points or foci as shown in Figure 19. There are two focal distances for each point on the curve.

BASIC DEFINITION: A SIMPLY DRAWN CURVE

An ellipse may be defined as the path of a point which moves so that the sum of the two different distances to the two foci is always the same. This property is a unique and interesting characteristic. The ellipse is a very easily drawn curve. All one needs is two pins, a pencil and a loop of string to continuously generate an ellipse. If one sticks the two pins into a piece of paper and then puts the end of the lead pencil inside the loop



$$\frac{\text{PF's}}{\text{All}} = \frac{\text{Focal Distances}}{\text{Directrix Distances}}$$

$= < 1 = \text{Eccentricity}$

$$= 0.66$$

$$\frac{x^2}{15^2} + \frac{y^2}{11^2} = 1$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

FIGURE 19

All (PF₁'s + PF₂'s) = AA₁ = 30

of string about the two pins (foci) and swings around, the lead pencil in the stretched out string draws out an ellipse. The loop of string has three parts; the one part is the constant length straight part between the two pins, (the two foci); another part of the string stretches from one focus to the pencil point and the third part of the string stretches between the other focus and the pencil point. As the pencil swings around the two pins, the sum of the latter two parts of the taut string is always the same distance, as is of course the distance between the two pins (foci).

AREA OF ELLIPSE

It is in order immediately to ask about the area of an ellipse in emulation of the Greeks and their excellent doctrine of always analysing geometrical problems from the point of view of areas.

The area of the ellipse evolves readily by comparison and analogy to the area of a circle

$$\text{Area of circle} = \pi r r = \pi r^2$$

$$\text{Area of ellipse} = \pi ab$$

The "a" is one half the length of the ellipse and "b" is one half the height or breadth. The "a" is one half the major axis and "b" is one half of the minor axis. It is not imperative but the major axis is usually thought of as being horizontal.

If the ellipse is "squeezed" so that "a" becomes equal to "b", a circle results and the area πab becomes πr^2 .

CIRCUMFERENCE OF AN ELLIPSE

The circumference of a circle is the diameter of the circle times (3.1416).

$$C = \pi d = 2 \pi r$$

This relation is a simple one and is the basis for a crude numerical evaluation of π . Nearly perfect cylinders can easily be made on a lathe. With such a cylinder, a piece of string and a ruler, the value of π can be approximated.

The π is an incommensurate number but it is about $3\frac{1}{7}$.

The circumference of an ellipse does not turn out to be so simple. However, it is closely evaluated numerically by the following formula.

$$C = 2\pi \frac{\sqrt{a^2 + b^2}}{2}$$

If $a = b$, the ellipse becomes a circle with radius (r). The formula for circumference then readily becomes $2\pi r$.

$$\text{SUM OF FOCAL DISTANCES} = \text{THE MAJOR AXIS} = 2a$$

If in the continuous tracing of the ellipse, the pencil point is at the extreme right of the ellipse and therefore on the X coordinate axis, the looping or moving part of the string will extend from the above extreme right point to the left passed the right focus and on to left focus and back again passed the right focus to the pencil point itself. The length of the looping part of the string is of course the sum of the focal distances as per one definition of the ellipse. In the above assumed position of the looping part of the string, the string consists of three portions. One portion is the distance between the foci, (the two pins). The other two portions are equal and are the distances of the foci to the near intersections of the ellipse with the X coordinate axis on the left and right side of the ellipse. It is readily seen that the length of the string is the major axis ($2a$) of the curve. The sum of the focal distances for all other points on the curve must likewise be the major axis or $2a$.

In terms of ellipse drawing, Figure 19, the focal distances point "P" are PF_2 and PF_1 . The distance PF_2 is $\sqrt{7^2 + 2^2}$ or 7.28. The distance PF_1 is $\sqrt{7^2 + 22^2}$ or 23.08. The sum is 30.36 which is a fair check for an actual tracing of an ellipse whose major axis is $2a$ or 30.

For point A_1 of drawing, the focal distances are A_1F_2 (5) and A_1F_1 (25). The sum is $2a$ or 30; the major axis A_1A of the continuously traced ellipse.

For a point at mid point of top of ellipse, the two focal distances are each the $\sqrt{11^2 + 10^2}$ or 14.9. The sum is 29.8; a fair check for an actual traced curve with a pencil, string and two pins.

In the later discussion and analysis of the properties of the hyperbola it will be pointed out that the difference, in contrast to the sum, in the two focal distances for all points on the hyperbola is equal $2a$. The " a " in this latter case is the X coordinate of the intersection of the hyperbola with the X axis.

A SECOND DEFINITION OF AN ELLIPSE

In addition to the first definition and specification of the ellipse previously given, a second simple specification can be written to prescribe the ellipse.

The ellipse is a curve or the path of a point which moves so that the ratio of its distance from a fixed point, called the focus, to its distance from a fixed line called the directrix, is always the same. However, the ration of the two distances must always be less than unity (1) in order to get an ellipse. The ratio of these distances is appropriately called the eccentricity, e . As said before, the eccentricity (e) of an ellipse is less than unity (1).

In the entire family of curves called the conic sections, there are four curves; the circle, the ellipse, the parabola and the hyperbola. They are called conic section because they all can be obtained by cutting a swath through a right circular cone at different angles. See charts and analysis of charts on pages 87, 90, 92, Figures 39, 40, 41.

The circle is a conic section which really has zero (0) eccentricity. The eccentricity (e) of the ellipse as above defined is less than unity (1). The eccentricity of the parabola is actually equal to unity (1), while the eccentricity of the hyperbola is greater than unity (1). So it is seen that the four conic section curves, (circle, ellipse, parabola and hyperbola) are of one family, are of like kind but different eccentricity. They are like four members of a human family with different characteristics, personalities and eccentricities even though the circle has no eccentricity. It might be better to say that the eccentricity of a circle is zero (a).

The interesting aspects of the various angles of intersection of a plane with the cone are discussed later. (Pages 87, 90, 92). Figs. 39, 40, 41.

As per Figure 19 and notations thereon, the first definition or specification for ellipse decrees that for any " P ", the sum of PF_2 and PF_1 is the same as for every other point on the curve, for example, P_1F_2 plus P_1F_1 . These two distances are not shown with lines on drawing but can readily be imagined.

The second definition or specification of the ellipse decrees that PF_1 is less than PD_p , where PD_p is the straight line distance from point P to the left directrix shown in drawing.

Also

$$P_1F_1 \text{ is less than } P_1D_{r_1}.$$

also

$$AF_1 \text{ is less than } AD_A$$

also

$$A_1F_1 \text{ is less than } A_1D_{A_1}$$

For all four points, P , P_1 , A and A_1

$$\frac{PF_1}{PD_r} = \frac{P_1F_1}{P_1D_{r_1}} = \frac{AF_1}{AD_A} = \frac{A_1F_1}{A_1D_{A_1}} = \text{eccentricity} = e = \text{less than unity (1)}.$$

For a point, P , the X distance is OM

$$\frac{OM^2}{OA^2} = \frac{\text{abscissa}^2}{\text{semi-major axis}^2} = \frac{X^2}{a^2}$$

$$\frac{PM^2}{BO^2} = \frac{\text{ordinate}^2}{\text{semi-minor axis}^2} = \frac{Y^2}{b^2}$$

For point P_1 , there would be an abscissa and an ordinate, and X_1 and a Y_1 . For P_1 we could have the same kind of an expression; namely $\frac{X_1^2}{a^2}$ and $\frac{Y_1^2}{b^2}$.

The basic equation of the ellipse is

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1$$

If "a" were equal to "b", the ellipse becomes a circle whose equation is

$$\frac{X^2}{R^2} + \frac{Y^2}{R^2} = 1$$

From point of view of trigonometry the above equation for the circle states

$$\sin^2 \theta + \cos^2 \theta = 1$$

There is no trigonometric equivalent equation for the ellipse.

Two Directrices

Since the ellipse has two focal points F_1 and F_2 , one could consider the eccentricity on the basis of ratio of distance of any point to either focus and the distance to a particular directrix of which the ellipse has two. If F_2 is the point of reference, then one must think of a second directrix to the right of ellipse. In this case, the eccentricity (e) would have the same numerical value which is less than unity (1).

$$\frac{PF_2}{PD_F} = \frac{P_1F_2}{P_1D_{F_1}} = \frac{AF_2}{AD_A} = \frac{A_1F_2}{A_1D_A} = e$$

This second directrix would be a distance of

$$\frac{a}{e} - a = a \left(\frac{1}{e} - 1 \right)$$

to the right of point A_1 by 7.5 divisions of length.

Since for the particular ellipse drawn, the e is $\frac{2}{3}$ then

$$a \left(\frac{1}{e} - 1 \right) = 15 \left(\frac{3}{2} - \frac{2}{3} \right) = 15 \cdot \frac{5}{6} = 7.5$$

EQUATION OF AN ELLIPSE

The basic specification of the ellipse provides, without too much work and maneuvering, the general equation for the ellipse already given.

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1.$$

where X is the abscissa of point, P , and Y is the ordinate of point, P .

XX and YY are the coordinate axes with origin "0" at center of ellipse: "a" is the semi X "diameter" or major axis and "b" is the semi Y "diameter" or minor axis of ellipse.

This equation reminds one of the equation of a circle. If "a" and "b" were equal then one would have equation of circle

$$X^2 + Y^2 = R^2 \text{ (R, being radius of circle)}$$

Here then for the ellipse, we have an equation that is similar in appearance to the equation of a circle.

ELLIPTICAL SINES AND COSINES

There is not a single constant distance in an ellipse like the radius in a circle which lends itself to ready use for possible analogy to circular sines and cosines. To be sure, there is the constant distance ($2ae$) between the two foci but the X and Y of a point on the ellipse do not form an angle with this distance.

However, it is neither necessary nor imperative that the constant distance reference line be a side of an actual angle to have a sine and cosine function. It could well be that an elliptical sine be $\frac{Y}{2ae}$ and an elliptical cosine be $\frac{X}{2ae}$. If this were done, nothing much would be gained. It would not provide any possible useful re-arranging of the fundamental equation of the ellipse into a trigonometric form. The radius of the circle played the role of a hypotenuse with constant numerical value available to serve as the denominator of sine and cosine ratios and does make an angle with the X and Y of a point on the circle. It is to be pointed out again that this fact of being a side of an angle is neither necessary nor imperative.

Using the circle, simple geometry, algebra and a ratio between ordinate and the single distance (radius), the "trig" man got the sine of angle. He also got a cosine. They are simple and readily interpretable ratios with graphical portrayals. But again, the radius was one side of an angle and was a constant distance when the point rotated counter clockwise in circumference round and round. Angles and sines and cosines came about in a simple and natural way in the case of a circle and provided the simple relation $\sin^2 \theta + \cos^2 \theta = 1$. It is a good practice to call these ratios not just sines and cosines but specify that they are circular sines and circular cosines.

In the ellipse drawn, again consider the point, P , as an example of any point in the curve. If a line were drawn from P to the center of the ellipse " O ", there would be an angle POM . At this point, the sine of the angle POM would be $\frac{PM}{OP}$. But when point P is moved around counterclockwise, the new distance P_1M_1 would not only change but also the new P_1O distance would differ. There is no constant distance to form angles like the radius (hypotenuse) of the circle associated with " P " as a basis of comparison, as it rotates in an ellipse as there was for a point rotating in a circle, namely the radius. But in the ellipse, the constant distance associated with the ordinates (PM 's) is another ordinate namely " b " not a PO . But from the standpoint of "trig", the distance " b " is not a part of the angle POM , so the conventional fine idea of a sine that easily and naturally came out of a circle is not possible now. In other words, as the point, P , rotates counterclockwise, various angles (POM 's) would of course alter the picture but we must abandon the fine and simple conventional idea of a sine. One might call the PM 's divided by the PO 's elliptical sines. However, since both PM 's and PO 's would be different for each point, the elliptical sine idea would not make sense.

Once again it is worth of observation, that $\frac{Y}{a}$ and also $\frac{X}{b}$ are ratios of two ordinates and two abscissas respectively. The " b " and the " a " are different in magnitude. The situation is not like the case of a circle when the X and Y were each compared to a single constant " R " which is the radius and which can be thought of as a hypotenuse. The two different distances " a " and " b " therefore play no part in the trigonometry of the angles in an ellipse in an attempt to have a basis for a geometrical representation of functions which might be called elliptical sines and cosines.

Even though the circular sine and cosine are ratios between an ordinate and a hypotenuse and an abscissa and a hypotenuse respectively there is no reason why we could not compare an ordinate to an ordinate and an abscissa to an abscissa to get a different kind of a sine and a cosine relation. This departure from the conventional procedure used for circular sines and cosines would be entirely legitimate and valid. In other words, the lines or distances do not need to be the sides of an actual angle to be compared so as to arrive at some kind of a ratio. It would be necessary that the ordinate and abscissa used for reference in the denominator of the ratio be the same magnitude. In the ellipse, the " a " and the " b " are not the same magnitude, hence elliptical sines and cosines are not logical.

One can look forward to the discussion of the idea that hyperbolic sines and cosines are arrived at on the basis of comparison of an ordinate to another ordinate and an abscissa to another abscissa, respectively. In this case of the hyperbola, the ordinate and abscissa used for reference as the denominator of the two ratios are equal. They have a different graphical place in the geometry of the hyperbola, to be sure, but they are equal in numerical value. They are not sides of an angle, however, which fact is neither necessary nor imperative as a general specification and definition for sine and cosine.

In a circle with a constant distance (radius) playing the role of a hypotenuse as a constant reference distance in a right-angled triangle, circular sines and circular cosines are a "natural." The equation, $\sin^2 \theta + \cos^2 \theta = 1$, for the circle is also a "natural."

CIRCULAR AND ELLIPTICAL ROTATION

Circular rotation is a common phenomenon. The wires in the armature of an electric generator and motor rotate in circles. The circular sine and the circular cosine curves previously shown in this treatise are good mathematical tools to put on the job and use in the study of alternating current phenomena. See later pages 147, 149, Figures 56, 57.

Assume one has a wagon wheel 4 feet in diameter which is spinning in a vertical plane on its axle. If, from some considerable distance away to the right, one looks edgewise at the spinning wheel (looks at the wheel in the plane of the wheel and at right angles to the axle) and notes a white spot on the outer edge of the wheel, the white spot will seem to move up and down (north and south) in a vertical line. It would move in what is commonly called simple harmonic motion. It would be a circular sine wave motion in other words as far as the vertical movement and vertical component itself was concerned.

If from a point way above the wheel, another observer noted the same spot, he would think the spot was merely moving to right and left (east and west). This apparent right and left straight line motion would also be "a la" simple harmonic motion.

Simple harmonic motion can simply be described for the present as circular motion projected on a diameter at right angles to line of sight of the observer.

Now however if an object or a spot were moving in an elliptical path (a lop sided wheel) and the object or spot were viewed sidewise or edgewise from a far east position like that mentioned before, the up and down (north and south) seemingly straight line motion along vertical diameter would be simple harmonic motion with amplitude equal to "b". The viewer, looking from the east would not know the object was moving in an ellipse. The distance up and down would each be equal to the vertical "diameter" of the ellipse.

If on the other hand, the ellipse were viewed from way up north as it were, the motion would again be simple harmonic motion (east and west) along the other "diameter" of the ellipse with an amplitude of "a".

One important point now to be made is that the motion projected along the minor axis and also that projected along the major axis are simple harmonic motions and provide recurrent, repetitive and periodic motion. The object, moving counterclockwise direction around the ellipse, does come back to the starting point. In this respect elliptical rotation is similar to circular rotation. Again periodic recurrence is a characteristic of bodies moving along the perimeter of a circle and an ellipse.

The S.H.M. motion projection along each axis is portrayed by its own individual sine and cosine curve.

The circle was described as the result of two simultaneous S.H.M. motions at right angles with equal amplitude and 90° difference in phase.

TWO S.H.M.'s AT RIGHT ANGLES

The ellipse is the result of two simultaneous S.H.M. motions at right angles with different amplitudes and like the circle having a phase difference of 90° degrees. The vertical S.H.M. component on Y axis is a sine curve of amplitude "b" while the horizontal S.H. component on X axis is portrayed by a cosine curve, 90° ahead in phase, with an amplitude of "a" along the other axis of the ellipse. Two viewers would think the object was moving in a S.H.M. fashion along a straight line with an amplitude of "a" in one case and in amplitude of "b" in the second case. Even though the path is an ellipse, the two S.H.M.'s at right angles function as though each of them were circular projections on the "a" and "b" axis. The frequency of each of these motions is obviously the same.

The simple harmonic motions at right angles, differing not only in phase but also in frequency, produce very interesting and beautiful curves called Lissajous curves like figure eights, etc. See page 49, Fig. 24.

CIRCULAR AND ELLIPTICAL PENDULUMS

At this point, it is suggested to the reader, that he go back to the first part of this memorandum about the circle and take a review look about 2 S.H.M.'s at right angles for a circle and an ellipse previously presented in the section on "Circles".

In one case, as previously discussed, it might be said that one has a circular pendulum and in the other case one has an elliptical pendulum. One pendulum "bob" supported by a string rotates in a circular path and the other pendulum rotates in an elliptical path.

Observers stationed at different points to the "east" and the "north" and looking at the same level as the plane of rotation would think they were ordinary pendulums swinging back and forth in a straight line. The two amplitudes at right angles for the circular pendulum would be equal. The two amplitudes at right angles for the elliptical pendulum would be unequal. See pages 30, 31, Figures 17, 18.

SUMMARY

Finally again, the idea of sines and cosines (even the unconventional idea of sines and cosines used in case of the hyperbola) can not apply to the ellipse whose equation does not involve a single distance like the radius of the circle to play the role of a hypotenuse of a triangle but rather two different distances. One of these distances is "a", one semi "diameter" of the ellipse and "b", the other semi "diameter".

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1$$

Here a abscissae (X 's) are compared to another fixed abscissa (a) and ordinates (Y 's) are referred to another fixed ordinate (b).

The "a" and the "b" do not enter the picture when the angles (POM 's) are considered; they are not hypotenuses; they are different in numerical value.

Pythagoras' theorem and the ideas of sine and cosine can make no contribution to elliptical paths in a fashion comparable or analogous to circular paths.

ELLIPTICAL PATHS

Most of the planets, including the earth, move around the sun in almost perfect circles. Actually the earth's path around the sun is an ellipse with not too great percentage difference in the major and minor "diameters". The eccentricity is quite small.

A fine example of an ellipse is the path of Halleys' comet. Halleys' comet was visible in America in 1910. It goes along its path out in space and will recur and come back in 76 years. It was seen years ago namely 1834 and noted by Halley 76 years before that in 1738. It was he who then calculated its elliptical path and its period of revolution of 76 years.

CHECKING DISTANCES ON ELLIPSE

It will be seen from drawing of ellipse (Figure 19), that "a" is 15 arbitrary units in length and "b" is 11. This ellipse was actually drawn by the method previously mentioned (a piece of string, two pins and a pencil). The distance from O to each focus (F) is "ae" or 10 units by actual measurement.

Since "ae" is 10 and "a" is 15, the eccentricity is $\frac{2}{3}$ which is as per specification namely less than unity (1).

As per specifications and as shown in second drawing, the directrix is a line parallel to Y axis and is at a distance of a/e to the left of origin of axes (O), namely a distance of 22.5 units.

With these actual distances at hand, we can make three appropriate calculations. Consider point, P , on curve; its abscissa (X) is 12 and its ordinate (Y) is 7. The distance from P to directrix is seen to be 34.5. The horizontal distance from P to the left principal focus is 22. ($10 + 12$). Therefore the diagonal distance of P to F is the $\sqrt{7^2 + 22^2} = \sqrt{533} = 23.1$ as per Pythagoras.

Now according to basic idea of ellipse the eccentricity (e) is the ratio of distance of any point to focus and distance to directrix. This ratio is 23.1 to 34.5 namely $\frac{2}{3}$, as per above statement; $\frac{ae}{a} = \frac{10}{15}$.

This calculation for any point "P" on curve would also come out $\frac{2}{3}$ for each and every point on ellipse.

Second Definition

The second definition of the ellipse was that the sum of the two focal distances is a constant.

The diagonal distance from P to left focus is as calculated above equal to 23.1.

The diagonal distance from P to right focus $\sqrt{7^2 + 2^2} = \sqrt{53} = 7.28$. The sum of the two focal distances is 30.38.

Now consider any second point on ellipse as P_1 , for example. Its two diagonal distances to the foci are $\sqrt{1^2 + 9^2} = \sqrt{82} = 9.06$ and $\sqrt{19^2 + 9^2} = \sqrt{442} = 21.02$. The sum of these is 30.08. The other above sum was 30.38. The check is pretty good considering the ellipse was actually drawn with two pins at the foci, a piece of thread and a pencil point which latter could not be held in absolutely vertical position as the "swoop" all around the foci was made.

Finally consider original point " P " again and the equation

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1$$

The coordinates of point, P , with respect to origin of axes at center of ellipse are 12 (the X distance) and 7 (the Y distance).

$$\frac{12^2}{15^2} + \frac{7^2}{11^2} = \frac{144}{225} + \frac{49}{121} = .64 + .40 = 1.04$$

This check for a "free hand" drawing of the ellipse curve is pretty good. Check for point P whose coordinates are 9 (Y) and -9 (x)

$$\frac{9^2}{15^2} + \frac{9^2}{11^2} = .36 + .67 = 1.03$$

Discussion and analysis of ellipse has afforded opportunity for careful thinking. From here, this treatise proceeds to the parabola and hyperbola.

CHAPTER 4

The Parabola

The circle and the ellipse discussed in the previous sections of this treatise are curves which prescribe the path of an object which might move along and around the same path again and again. Since they are closed curves, the moving object or the mathematical function as the term is used in such matters has a periodic and recurrent aspect in its career. The circle and ellipse are called "conic section curves because they are associated geometrically with the intersection of a plane and a right circular cone at different angles of intersection.

Another conic section curve is the parabola. The reader will like to look at the photographs at page (8) Figures 1-13, which portray the various intersections to provide a circle, ellipse, parabola and hyperbola. Before even stating the special character of a parabola, it is to be pointed out and emphasized at the very outset that the parabola is not a closed curve. This is a very important point to keep in mind, as the conic section curves relate in general to physical and electrical phenomena. There can be no periodic or repeated journeys when an object moves in a parabolic path. Hypothetically, one might say that an object might journey out to infinity and then come back some time. Such an observation is only theoretical and supposititious. The object goes "off into the wide blue yonder" never to return.

NATURE OF PARABOLA

A parabola is the path or locus along which an object moves so that its distance from a fixed point called the focus is always numerically equal to its perpendicular distance to a fixed straight line. This latter line is called the directrix. As stated earlier in the previous section, the ratio of these two distances in this case is unity (1) and is called the eccentricity, e .

In terms of accompanying drawing, Figure 20, the above specifications for a parabola mean that, for any point P on any of three curves, the distance $PF = PD$. The PD is the distance to the directrix. For any and every other point along the curve its PF will equal PD . The reader might like to measure, with a divider and ruler, the two distances for several points on each parabola and note that the two distances in each curve are equal. The PF distance is called polar radius. This polar distance is of course a different distance for every point and gets bigger and bigger as the object moves out along its path. The parabola is not like the circle whose every point is equi-distance to the center of the circle.

If one does some simple algebraic and geometric maneuvering in terms of the basic specifications for the parabola ($PF = PD$; $P_1F = PD_1$) one readily comes out with a simple equation for the parabola when referred to the coordinate axes of analytical geometry.

The equation for a parabola is $Y^2 = 4pX$

It is interesting and important to observe that unlike the circle and the ellipse already discussed and the hyperbola to be subsequently discussed, the X (abscissa) in the equation of the parabola is not squared (X^2) but is just X .

This equation as interpreted with the help of the above sketch says that the square of the ordinate PM (Y) for any and every point on the curve is equal to four times the product of the abscissa $OM(X)$ and the constant distance (p) from the focus F to the origin of axes, point O .

One might point out that the "directrix", previously identified as the line to which the distance (PL) the point P was referred to, lies to the left and parallel to Y axis. This directrix line is up and down at a distance of " p " to the left of Y axis.

As per specifications for the parabola, the reader might be interested to observe that for point H , the distance to directrix is $2p$ and the distance to F is also $2p$. For the point O which is not only the origin of coordinate

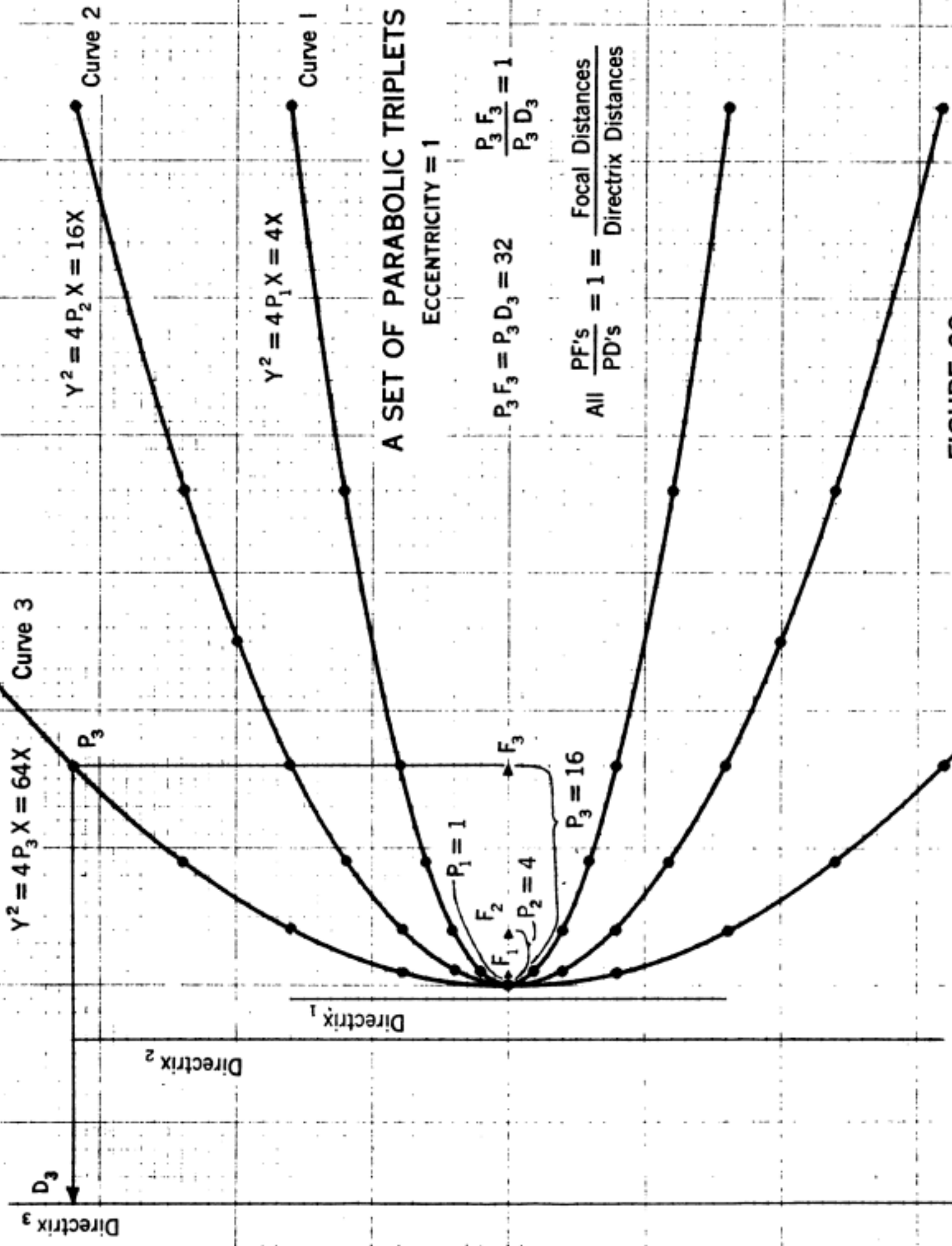


FIGURE 20

axes but also a point on the parabola, one distance (to directrix) is p and X the other distance to focus is also p . The ratio of distances for both points is unity; therefore, $e = 1$.

NO PARABOLIC SINES AND COSINES

As previously pointed, the conventional circular sines and cosines indicate recurrency and periodicity. Since the parabola is an open (not closed) curve, it is of no avail to try to apply the conventional circular sine and cosine idea as they apply to angles in a circle.

But just for fun, say we do make a try. Assume we draw a straight diagonal line from P to O . (not actually drawn in sketch)) Then we would have a nice right triangle POM and a nice acute angle at O , namely angle POM . This line PO might be a hypotenuse for Pythagoras. It is true that $Y^2 + X^2 = PO^2$ but point P is not moving around O in a circle at a constant radius of PO^2 but point P is not moving around O in a circle at a constant radius of PO . The point P is embarked on a journey out along a parabola; "off it goes into the wide blue yonder". When the point is way out, it is again true that its $X^2 + Y^2$ would equal its PO^2 but now it is a different PO .

The $X^2 + Y^2 = PO^2$ would always be changing unlike the $X^2 + Y^2$ of a circle which for any point in the circular path is always equal to a constant radius².

So it is of no avail to talk either about the sine of angle POM being PM (Y) divided by PO nor about the cosine of angle POM being OM (X) divided by PO because our attempt does not supply any useful information about the route along which the object is moving.

If we would want to know anything about the route to "the wide blue yonder", we can only say that wherever we are

$$Y^2 = 4PX$$

PARABOLIC SINES AND COSINES

In an attempt to utilize and take advantage of the sine and cosine general idea, we might argue that a new approach and concept of sine and cosine might be invoked in connection with the parabola. We made the same attempt in connection with the ellipse.

For the circle there was the single constant distance (hypotenuse) available to which the X 's and Y 's could be compared. In the parabola, there is the constant distance, $2p$ which is one half the entire chord $4p$ or HK in the drawing. Then there is also the constant distance " P " from O to the focus (F), namely OF in the sketch.

We might amend and revise our conventional idea of a circular sine and say that the parabolic sine of angle POM is Y (an ordinate) referred another ordinate P , one half HK . We might say that cosine of POM is X (an abscissa) referred to another equal abscissa, P . (OF)

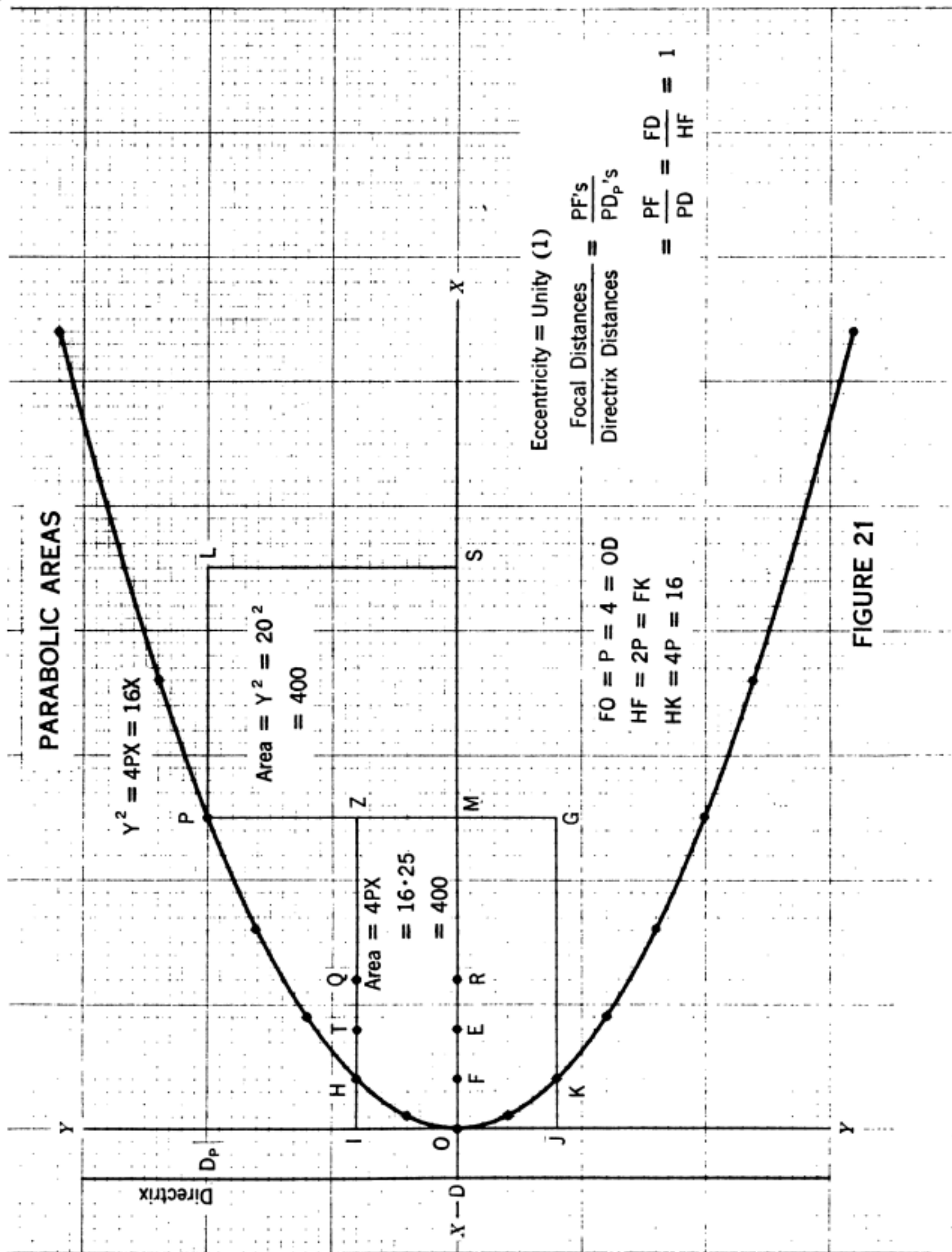
It is likely that this procedure would be of little avail and offer no particular useful contribution to the problem of quantities like voltage and current that might vary in parabolic fashion in electrical engineering in general and communication engineering in particular.

Since the X in the equation for the parabola ($Y^2 = 4PX$) is not squared as it is in the circle and hyperbola, there would be no likelihood of effectively using the sine² plus cosine² idea in the equation for interpretive purposes.

However, as stated in connection with the ellipse, it is interesting to note that in connection with the hyperbola and hyperbolic sines and cosines to be discussed later, the above suggested idea of comparing X and Y distances (abscissas and ordinates) to another abscissa and another ordinate respectively is actually used. Of course the abscissa and ordinate, to which the X and Y of the hyperbola are compared respectively, are equal in magnitude.

AREA UNDER PARABOLA

Consider any point " P " Figure 21, with coordinates X, Y on the parabola $Y^2 = 4PX$ in accompanying chart. The point " P " has an ordinate of length Y and an abscissa of length X . If one imagines a rectangle built about the parabolic sections with the base X (OM) and height, Y , (PM) the area of the rectangle would be XY . The



parabolic area within the rectangle would be bounded by X (OM) on the bottom, by Y (PM) on the right and the curve itself. (OP) In keeping with the doctrine of looking at matters from the standpoint of areas, it is interesting to note that the parabolic area ($PMOP$) is $\frac{2}{3}$ of the rectangular area.

This ratio of $\frac{2}{3}$ would apply to the two areas involved for any point on any parabola. This relation is supplied by a splendid "assist" from integral or "adding up" calculus.

PARABOLIC PATH, A COMBINATION OF TWO MOTIONS OFTEN ENCOUNTERED IN EVERYDAY LIFE

The circle was identified as the combination of two simultaneous simple harmonic motions at right angles with equal amplitude and a phase difference of 90° . The ellipse was identified by an exactly similar specification except that the amplitudes were different.

The sine curve is the combination of one S.H.M. (on Y axis) and uniform velocity at right angles (on the X axis).

An excellent way to regard a parabola is that it is, like the sine curve, the circle and ellipse, the combination of two motions at right angles.

A DRINKING FOUNTAIN PARABOLA

In fact, the parabolic path is seen in our natural physical environment almost as frequently as the circular paths of points on spinning wheels of all kinds.

The first motion to produce a parabola path is uniform velocity in one direction while the other motion at right angles is uniformly accelerated motion. Uniformly accelerated motion is very common since "all things which go up must come down" and they do so with uniformly decelerated motion and uniformly accelerated motion respectively. This means that a projectile like a bullet from a gun, the bomb dropped from an airplane or the "shot" as put by a "shot" putter in an athletic field meet has a parabolic path. The water stream out of a hose has a parabolic shape. Oftentimes large fountains in public parks present a beautiful ensemble of parabolas.

The uniformly accelerated motion is that provided by the force of gravitational attraction which provides a deceleration of 32 feet per second per second as things "go up" and an acceleration of the same amount as things "come down". In such cases, one gets a parabola whose main axis is vertical. This means that the bullet or water stream projected horizontally is a half parabola whose principal axis is vertical.

The accompanying picture Figure 22 of a drinking fountain water stream shows it to be a beautiful parabola. If water emerges from a series of holes in a circular ring as in some types of a lawn sprinkler, one sees a beautiful array of family of slender streams of water each of which is a parabola.

In a recent newspaper account of a base ball game it was written; "the batter caught a 2 and 2 pitch and drove it in a screeching 400 foot parabola over the rightfield fence". The batter had supplied the uniform velocities and gravity had supplied the uniformly decelerated velocity on the way up and the uniformly accelerated velocity on the way down.

Every pitch, every ball thrown, bounced or batted in a baseball game has a parabolic flight. The many parabolas have different shapes with different mathematical constants. Even the swiftest Walter Johnson pitch executed a parabola. All games, tennis, golf, football etc., in which a ball in flight is involved are parabolic games.

There was an arithmetical progression in distance on the way up of 32 feet during each individual second and the same arithmetical decrease on the way down. Likewise there was an arithmetical increase and decrease in velocity during each second of 32 feet per second on the way up and on the way down respectively.

DRAWING A PARABOLA

The easiest way to draw a parabola is to use the idea of previous paragraph involving a combination of uniform velocity up and down along the Y axis and uniformly accelerated motion on the X axis to the right of origin.



A DRINKING FOUNTAIN PARABOLA
FIGURE 22

2

ALPHABETIC LISTING OF NAMES &
THEir RESIDENCES



To do this latter, one could readily apply the two simple formula involving distance and time.

$$\text{Distance} = \text{Uniform Velocity times Time}$$

$$\text{Distance} = \frac{1}{2} \text{ Acceleration times Time}^2$$

These formulas would give distance that the object moves under uniformly accelerated motion out to the right on the X axis during the same period of time (t) involved in the up and down uniform velocity distance on the Y axis from the origin O .

Let us assume a uniform velocity up and down along the vertical (Y) axis of 50 feet per second. Assume really two moving objects starting from the origin of axis (O). One object moves up and the other object moves down. Now also assume each object moves simultaneously, with its up or down motion, with a motion to the right (uniformly accelerated) of 20 feet per second per second. The two objects will nicely plot two paths to form the two symmetrical parts of a parabola with the X axis as the main axis of the parabola.

Time in Seconds	Uniform Velocity (Y) Arithmetic Ratio Distance = $S = vt$	Uniformly Accel (X) Geometric Ratio Distance $S = \frac{1}{2}at^2$	Time ²
$\frac{1}{2}$ Second	25 feet	2.5 feet	$\frac{1}{4}$
1 Second	50 feet	10. feet	1
2 Second	100 feet	40. feet	4
3 Second	150 feet	90. feet	9
4 Second	200 feet	160. feet	16
5 Second	250 feet	250. feet	25
6 Second	300 feet	360 feet	36

The parabola shown in accompanying chart Figure 23 was plotted according to above easily calculated data.

The parabola may be (Missile Trajectories; Parabolic Arch) very well thought of in terms of two motions at right angles under the rules that the distances along one coordinate axis vary directly proportional to the time while the distances along the other axis vary directly proportional to the square of the time.

TWO POINTS OF VIEW

It is interesting to consider the different points of view that the physicist and the analytical geometrician have regarding the equations, $S = \frac{1}{2} aT^2$. To the physicist, this is the simple equation which provides the total distance (S) which a free falling body falls when the constant force of gravity impels it to move with uniformly accelerated velocity. The " a " is the acceleration in feet per second per second (32) and T is the elapsed time in seconds.

To the geometrician, $S = \frac{1}{2} aT^2$ is merely the equation of a curve, namely a parabola. The general equation of the parabola is $Y^2 = 4 PX = KX$. So the mathematician writes the falling body equation, $T^2 = KX$. If the time (T) is regarded as the independent variable to be plotted on one of the coordinate axes (the X axis), then Y , the dependent variable (to be plotted on Y axis) is directly proportional to the square of T . If one plots these two variables, a parabola results. So to the geometrician, $S = \frac{1}{2} aT^2$ is the equation of a parabola. The " a " or the " K " may have different values, but nevertheless $S = \frac{1}{2} aT^2$ is intrinsically and essentially, the equation of a parabola.

ARITHMETICAL PROGRESSION

Another interesting and significant way to describe the two motions on Y axis and X axis which produce a parabola is now presented.

CONSTANT VELOCITY

On the horizontal (X) axis for both up and down motion, the distance per unit of time was constant. This type of motion, with displacement per unit time a constant, is characteristic of uniform velocity. In each individual second, the displacement or distance was the same as for every other individual second.

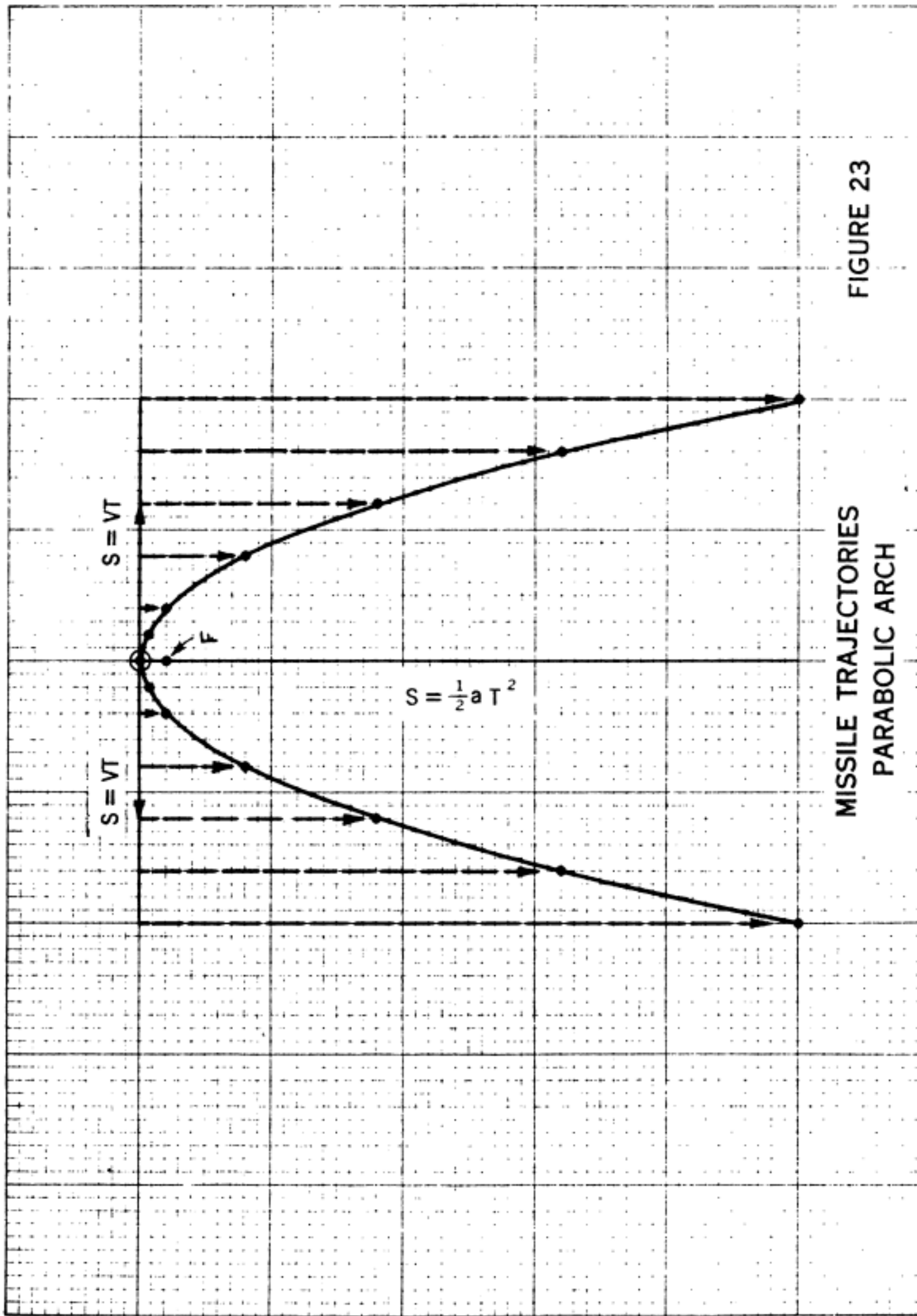


FIGURE 23

MISSILE TRAJECTORIES
PARABOLIC ARCH

VELOCITY INCREASES: ACCELERATED MOTION

On the other hand, the velocity for the vertical Y axis motion was not constant but was increasing at a constant rate of 32 feet per second during each succeeding second. The acceleration was 32 feet per second per second. The velocity increases by a constant amount or constant increment. The velocity increases therefore in arithmetical progression as time "marches on".

DISTANCE PER UNIT TIME INCREASES

As an accompanying result of the arithmetical progression of velocity, the X distance per individual unit of time also increases by a constant amount as time "marches on". The X distance per individual second therefore also increases in arithmetical progression on the X axis.

During each individual second, the distance traveled on X axis is 32 feet farther than the distance traveled in the previous individual second. In the very first second, the distance is 16 feet.

The Y distance per individual second is constant. The X distance per individual second increases in arithmetic progression. This is a distinctive characteristic of the uniformly accelerated motion which is happening on X axis. If in some case, one had uniformly decelerated motion, the individual velocities and distances for each individual time interval would decrease in arithmetic retrogression.

One can readily imagine an arithmetic retrogression in some circumstance in which case each term is less by a constant amount than the previous term.

In the case of uniformly accelerated motion, the increment of velocity and distance is constant and in the case of uniformly decelerated motion, the decrement is constant.

GEOMETRIC PROGRESSION AND RETROGRESSION

Geometric progression and retrogression, as contrasted to arithmetic progression and retrogression, involve a succession or series of quantities between which there is constant ratio. Each term in the series is for example 10 per cent greater or 10 per cent less than the previous term. Each term might be twice as great or $\frac{1}{2}$ as great, giving rise to a geometric progression and retrogression respectively. If the constant ratio, which can be any magnitude, is greater than unity (1), one has a geometric progression. If the constant ratio is less than unity (1), one has a geometric retrogression.

GEOMETRIC PROGRESSION AND RETROGRESSION FREQUENTLY MET IN NATURAL PHENOMENA

Many changes or variation in one of the factors involved in many natural phenomena occur in geometric progression and geometric retrogression fashion, as controlled by or dependent on some other factor. It seems that in natural phenomena, there are more decays or thinnings out or attenuations than growths or increases. In other words, it seems that geometric retrogression is more common than geometric progression. If one puts money in a bank at compound interest, one does have a good example of geometric progression. In a telephone circuit, the electric power embodying the voice decreases as distance increases arithmetically in geometric retrogression fashion. See page 125, Fig. 50.

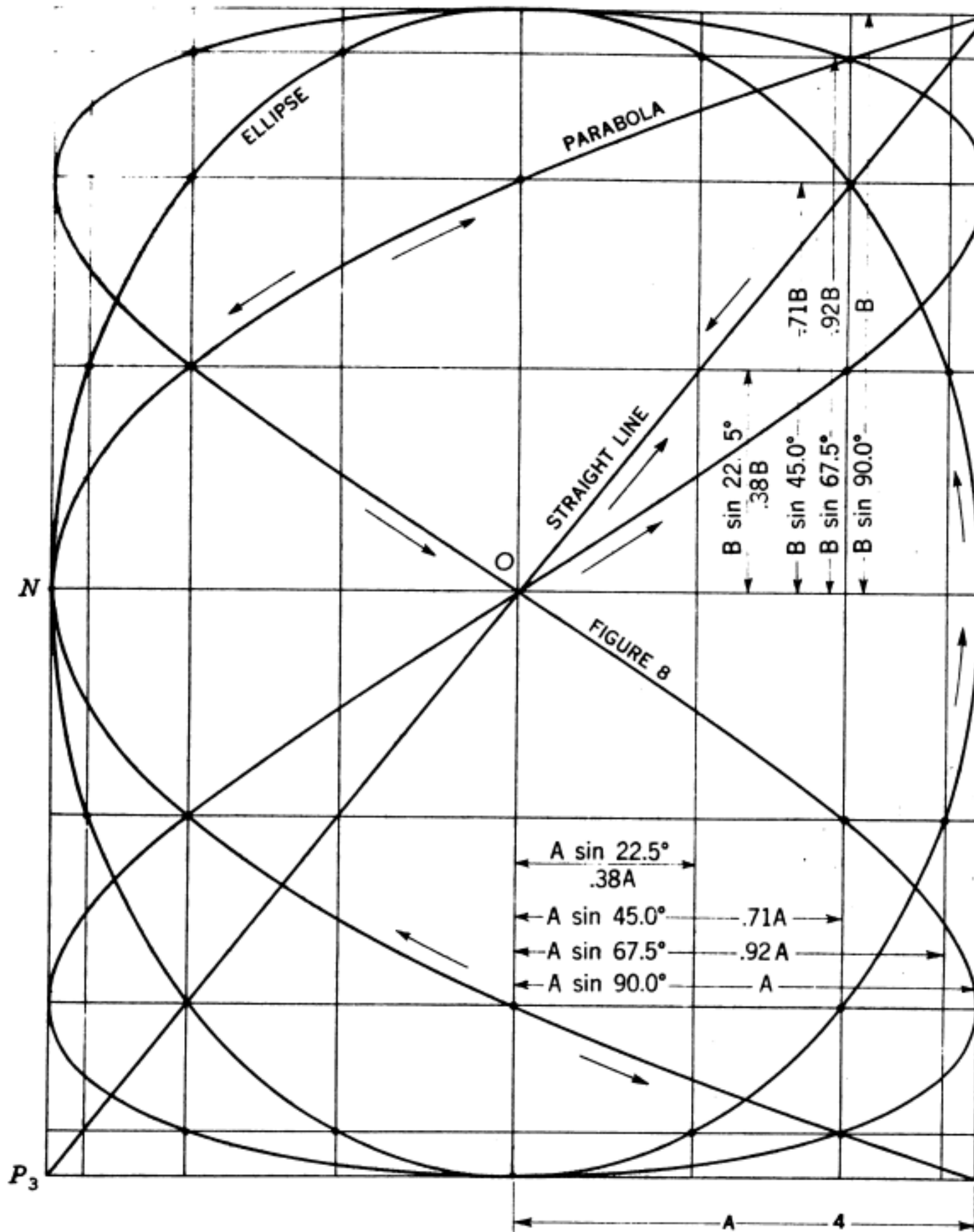
In a later section of this treatise, there is going to be considerable provocative discussion and illustrations of geometric progression and retrogression, particularly the latter.

FURTHER CONSIDERATION OF TWO SIMPLE HARMONIC MOTIONS AT RIGHT ANGLES

THE PARABOLA

In the discussion of the circle and ellipse and two simple harmonic motions at right angles at pages 30 and 31, Figs. 17, 18, of circle section of manuscript it was mentioned, that a parabola may be traced by a point controlled by two simple harmonic motions at right angles when the frequencies are two to one and the phase difference is 90° or 270° . See accompanying Figure 24.

This fact is particularly interesting and a bit surprising when contrasted to the fact, as previously discussed that the parabola is also the path of a particle moving with uniform velocity in one direction and uniformly accelerated velocity (arithmetical progression) at the right angled direction.



FREQUENCIES : 1 TO 1 : 2 TO 1

AMPLITUDES : 4 TO 5

: A TO B

TWO SIMPLE HARMONIC MOTIONS
AT RIGHT ANGLES

FIGURE 24

As previously discussed the mathematical equation of a parabola is in its simplified form, $Y^2 = 4PX$. The niceties and particulars of the mathematical proof that the curve in the accompanying chart is really a parabola and that it has an equation of the form $Y^2 = 4PX$ can be found by reference to text books in analytical mechanics. The figure 8 curve has an equation which involves the fourth power rather than the square.

Discussion and exposition of this latter equation is not in order here in the light of the objectives and scope of this treatise.

A look and discussion of the accompanying drawing, provides further interest in and understanding of two S.H.M.'s at right angles.

Analysis of Accompanying Chart, Figure 24

The several distances along X axis are the orthogonal or perpendicular projection on the X axis of uniform angular velocity (w) in a circle whose diameter is 4 inches. The several distances along the Y axis are the orthogonal or perpendicular projections on the Y axis of uniform angular velocity in a circle whose diameter is 5 inches. These circles are not shown in drawing but they can be imagined readily. Of course, the linear velocities along the two separate circular paths are equal and uniform.

The several individual distances on both X and Y diameters, starting at origin (0), are directly proportional to the sine of the angle $\theta(wt)$ in steps of 22.5° $\left(\frac{\pi}{8} \text{ radians}\right)$. This means that in the S.H.M. on each diameter, the several individual distances are of unequal value and are, it is interesting to note, the S.H.M. distances traversed in equal intervals of time, $\frac{T}{16}$. T is the time of a complete revolution.

THE ELLIPSE

Frequencies, 1 to 1: Amplitudes, 5 to 4: Phase Difference, 90° or 270° .

The ellipse in drawing, is generated by starting at point, M . The frequencies along the two axes are equal. The angular velocities are equal; $w = 2\pi f$.

By starting at point M , the Y motion lags the X motion by 90° $\left(\frac{\pi}{2}\right)$. If, from point M , one goes up one of the several individual different distances and also simultaneously goes left one distance, the first point on an ellipse is provided. Then one goes up an additional individual distance and also another distance to the left to get second point on the ellipse. By continuing to go up one distance and later down while one goes to the left one distance and later to the right, all points on the ellipse are provided. The tracing of the ellipse has been counter clockwise. If one had started at N with 270° difference in phase and at the start go up and to the right one distance at a time, the tracing would be in clockwise direction.

If the amplitudes of the two S.H.M.'s were equal, one would trace out a circle as per chart page 27, Figure 18.

STRAIGHT LINES

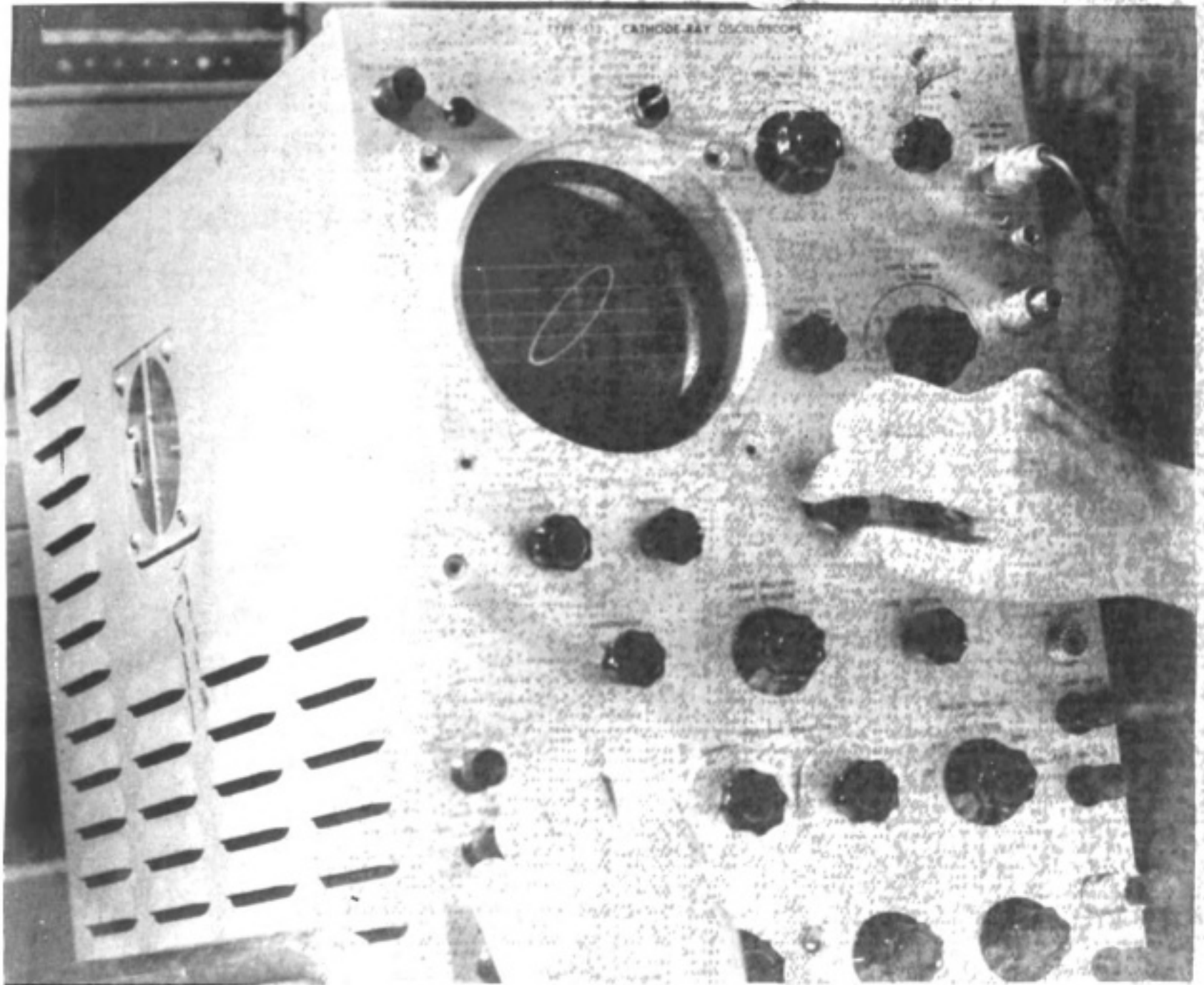
Frequencies, 1 to 1: Amplitudes, 5 to 4: Phase Difference, 0° or 180° .

If the start were made at origin (O), there would be no phase difference. By moving to the right one distance and simultaneously up one distance on the axes, one would get a northeast and south west straight line (OP_1OP_3). If one started with 180° difference in phase (Y motion lagging X motion) one gets a northwest southeast straight line.

A FIGURE 8 CURVE

Frequencies, 2 to 1: Amplitudes, 5 to 4: Phase Difference, 0° or 180° .

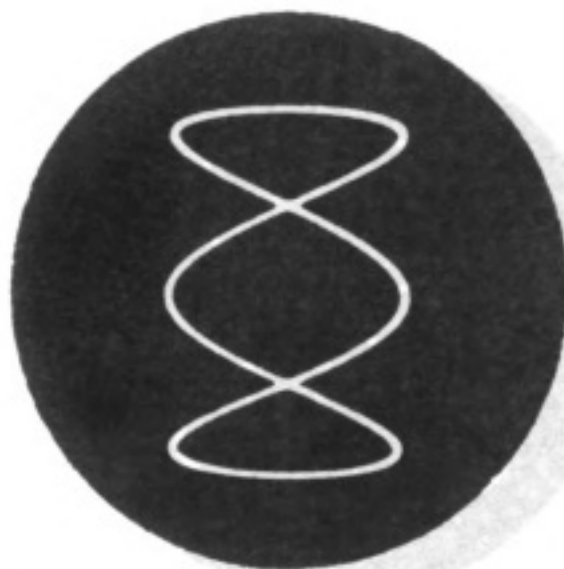
If one starts at origin (O), (no difference in phase) and moves two distances to right and later to the left while simultaneously moving one distance up and later down, one obtains a sort of figure 8 curve by the combination of two S.H.M.'s. If phase difference were 180° , the figure 8 is generated in a direction opposite to first case.



ELLIPSE OUTLINED
ON
CATHODE RAY OSCILLOSCOPE
FIGURE 25



FREQUENCIES 1 TO 1 PHASE DIFFERENCE 90° ELLIPSE



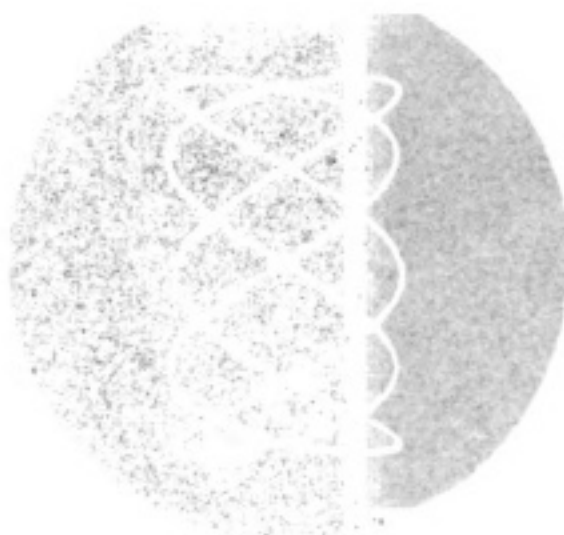
FREQUENCIES 3 TO 1 PHASE DIFFERENCE 90°



FREQUENCIES 5 TO 2 PHASE DIFFERENCE 0°

TWO SIMPLE HARMONIC VOLTAGES AT RIGHT ANGLES
CATHODE RAY TUBE PHOTOGRAPHS

FIGURE 26

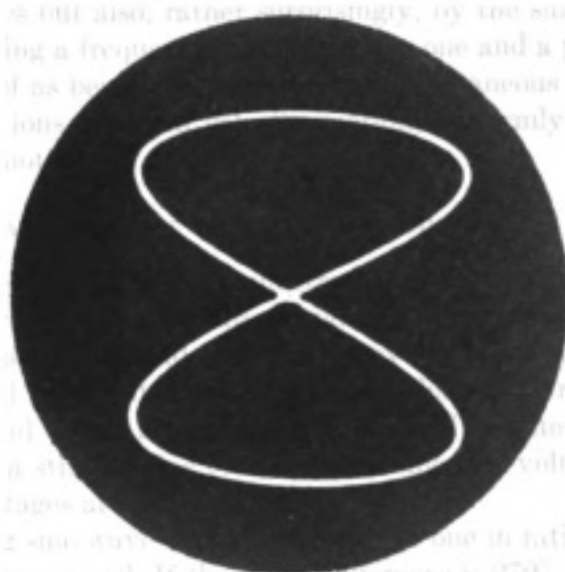




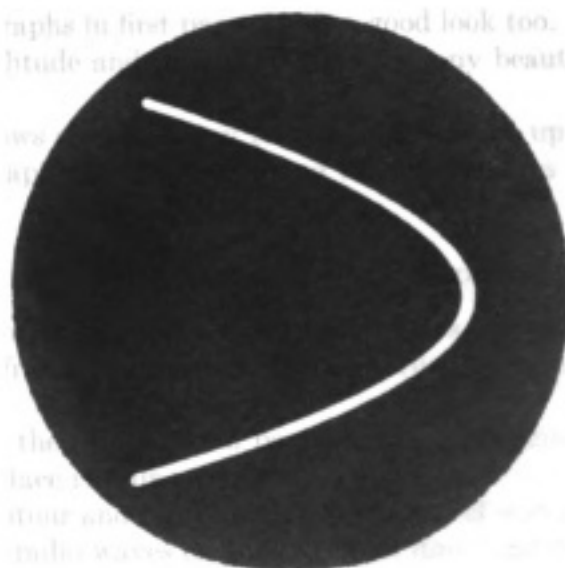
FREQUENCIES 2 TO 1

PHASE DIFFERENCE 270°

PARABOLA



FREQUENCIES 2 TO 1

PHASE DIFFERENCE 0° 

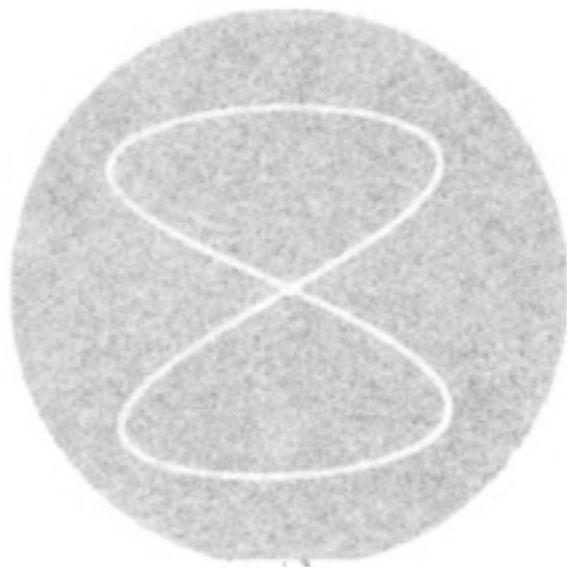
FREQUENCIES 2 TO 1

PHASE DIFFERENCE 90°

PARABOLA

TWO SIMPLE HARMONIC VOLTAGES AT RIGHT ANGLES
CATHODE RAY TUBE PHOTOGRAPHS

FIGURE 27



PARABOLA

Frequencies, 2 to 1: Amplitudes, 5 to 4: Phase Difference, 270° or 90° .

If one begins at point N , which means a phase difference of 270° (Y motion lagging X motion) and moves two distances to right and later to left and simultaneously moves up, and later down, one distance, a parabola NP_1NP_2 (facing to the right) is generated. If one had started at M with Y motion lagging X motion by 90° and moves two distances to the left on X axis (and later right) while moving simultaneously one distance up (and later down), then a left facing parabola is generated.

In contrast to the ellipse, circle and figure 8, in which case the moving point goes round and round without stopping, the point generating the straight line and the parabola goes to a certain position and stops. It then turns around and goes backward along the path it has just traced.

An idea previously mentioned regarding the parabola merits repetition.

The parabola is generated not only by the simultaneous combination of uniform velocity and uniform accelerated or decelerated at right angles but also, rather surprisingly, by the simultaneous combination of two harmonic motions at right angles having a frequency ratio of two to one and a phase difference of 90° or 270° .

The hyperbola cannot be thought of as being generated by the simultaneous combination of any two of the three sort of standard or natural motions; uniform velocity motion, uniformly accelerated or decelerated velocity motions and simple harmonic motion.

CATHODE RAY OSCILLOSCOPE CURVES

The ideas previously discussed pages 28, 29 are nicely illustrated by the application of two sine wave voltages to the two sets of electron beam control plates which are at right angles in certain types of cathode ray oscilloscopes. Page 52, Figures 26, 27. If phase, magnitude and frequency are properly adjusted, the spot of light, on the oscilloscope fluorescent screen will trace out a circle, ellipse or diagonal straight line for equal frequency. If phase difference is 90° the circle and ellipse are obtained with equal and unequal values of voltage respectively. If phase difference is zero (0) a straight line of light is obtained. If voltages are equal, the line will be at 45° and at some other angle if voltages are unequal.

If the frequencies of the alternating sine wave voltages are two to one in ratio and if the phase difference is zero (0), then a figure 8 path will be generated. If the phase difference is 270° , then a parabolic path () will be swept out by the controlled spot of light on the fluorescent screen.

If the phase difference is 90° the parabola faces the other direction ().

For further consideration of the parabolic path the reader might take a look at latter part of discussion of the parabola on page 44. The photographs in first pages merit a good look too.

Other combinations of phase, amplitude and frequency provide many beautiful and intricate curves called Lissajous's curves.

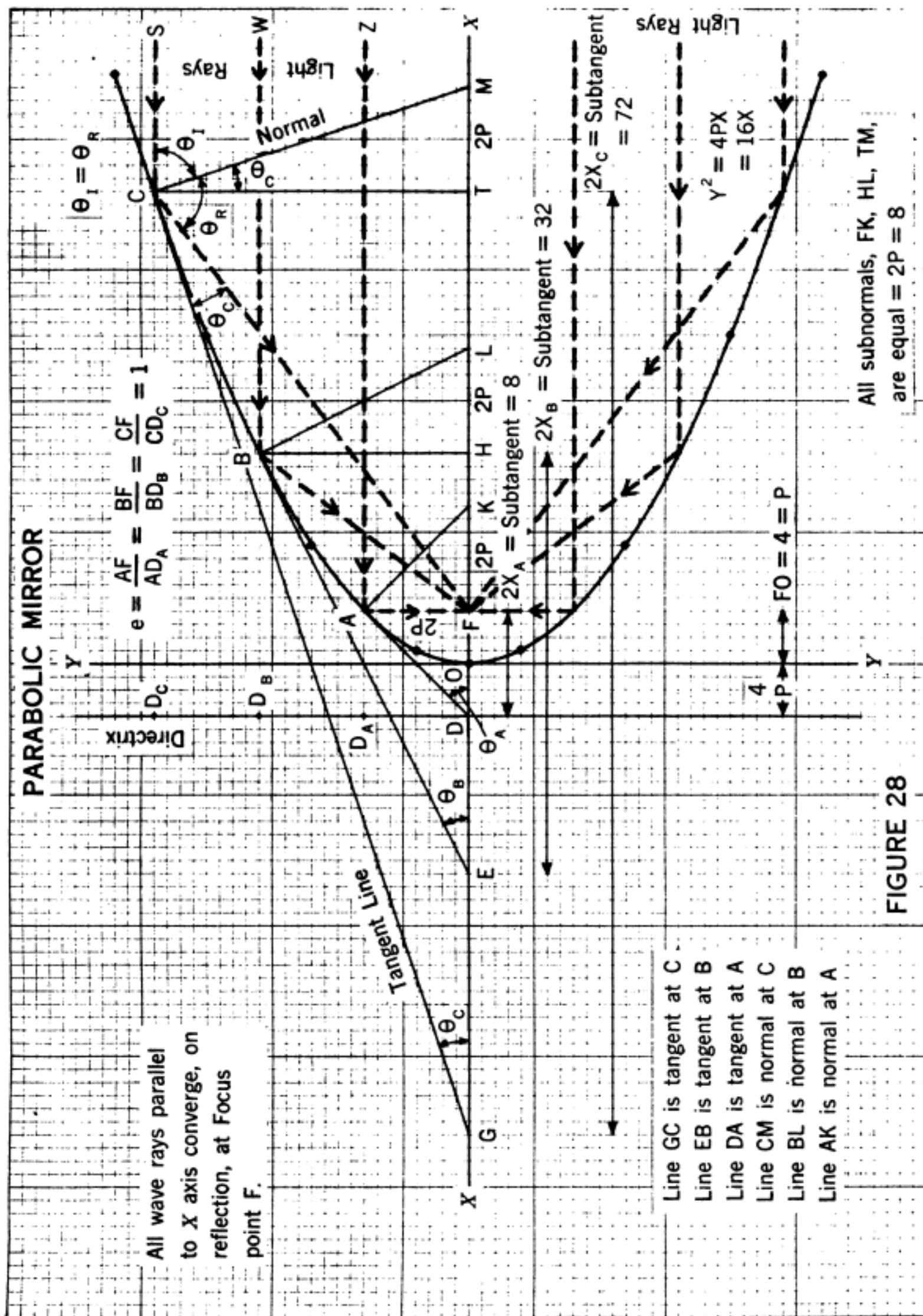
The accompanying photograph shows a cathode ray tube oscilloscope set up in a research laboratory with an ellipse on the screen. The photographs of the figure 8 and the other curves were obtained by the same apparatus.

A PRACTICAL USE OF THE PARABOLA

The paraboloid mirror is a fine example of the practical utility of this curve. It has a definite advantage over a spherical mirror in reflecting light from a source placed at its focus in a precisely accurate beam parallel to principal axis.

The parabolic mirror is formed by the rotation of a parabola about its principal axis. A three dimensional paraboloid with an inner concave surface is thus provided.

The parabola has just the right contour and right decreasing change of curvature outgoing along its path so that the many rays of light or micro radio waves emanating in all directions from an energy source placed at the focus, will all reflect parallel to the principal axis. Conversely, when many rays of light or radio waves parallel to the principal axis (coming from some distant energy source) reflect from the mirror surface, all will converge nicely and accurately at the focus. This same accuracy of focus is not characteristic of a mirror surface which is a part of a sphere (a "circleoid"). This inaccuracy of focus is called spherical aberration.



The many rays of light, emanating in many directions from a source of light at the focus of a paraboloid, strike its surface at different angles. As one can observe, by taking a look at a parabola, the curve becomes less steep as referred to the axis when one goes out on the curve. This means that the tangent line drawn from some point to the left on the x-axis to the point at issue on the curve, becomes less steep as the point of reflection involved is assumed to be farther and farther out along the curve. So it results nicely that all the many reflecting rays of light, going backwards, upwards, will hit at the proper angle of incidence so that the ever equal angle of reflection gets on the job to decree that all the outgoing reflected rays will be parallel to the principal axis.

MATHEMATICAL ANALYSIS OF THE FOCUSING CHARACTERISTICS OF A PARABOLIC MIRROR

The niceties and particulars of the proof of the previously stated focusing characteristics of a parabolic mirror are most interesting and intriguing. The proof merits exposition at this juncture as an important and final step in the consideration of the inherent nature of the parabola.

It is suggested to the reader that he take a good look at the accompanying drawing Figure 28 before further reading and study is done.

In the proof, the basic concepts of slope, tangent line at a point on a curve, perpendicular or normal line to the tangent line at a point on a curve, circular trigonometric function called the tangent (ratio of ordinate to abscissa) and similar triangles will be invoked. Furthermore, the simple fact or law of light reflection, which states that the angle of incidence is equal to angle of reflection, also will be called upon.

The first step in the proof will capitalize on the splendid contribution of that branch of mathematics called *differential calculus* as it applies to angle of slope of a curve at a point on the curve. The reader may not have studied and learned about the easy essentials of calculus. This fact need not stymie him in his commendable desire to appreciate and understand the main arguments of the proof which provide a simple and important conclusion.

The fundamental and basic expression of the differential calculus is $\frac{dY}{dX}$. This is actually the ratio of a little bit or a very tiny increment of Y to a little bit or a very tiny increment of X . It is an expression for the trigonometric tangent of the angle of slope of a curve at a particular point on a curve. It gives also the ratio between the ordinate Y of a particular point on a curve to the abscissa X of that point on the curve. The slope of a curve (or a straight line, too) at a point is the ratio of its ordinate to the abscissa of the curve at that point. Again it is to be said that $\frac{dY}{dX}$ is the trigonometric tangent of the angle of slope.

The equation of the parabola is $Y^2 = 4PX$. This can be written as $Y = 2\sqrt{P}\sqrt{X} = 2\sqrt{PX}^{\frac{1}{2}}$. We will now accept as a splendid "assist" (as per the baseball term) on the part of differential calculus the fact that if $Y = 2\sqrt{PX}^{\frac{1}{2}}$ then

$$\frac{d}{dX} 2\sqrt{PX}^{\frac{1}{2}} = \frac{2\sqrt{PX}^{-\frac{1}{2}}}{2} = \frac{\sqrt{P}}{\sqrt{X}}$$

This expression for $\frac{dY}{dX}$ is arrived at in accordance with the basic, simple and beautiful idea that

$$\frac{dY}{dX} \quad \text{of} \quad Y = CX^N = NCX^{N-1}$$

The C is a constant in general; it is $2\sqrt{P}$ in the case of the parabola.

The physicist and the engineer are not like the soldiers in "The Charge of the Light Brigade" who, under military orders, were "not to reason why." The physicist and the engineer like to and do "reason why" but, after being helped to understand and know "why" by the mathematician, they accept and make use of the doctrines and facts of mathematics as such. They gratefully and gladly give an accolade of praise and thanks to their talented colleagues.

The above expression for the trigonometric tangent of the angle of slope at any point along a curve means,

in mirror drawing, that $\frac{dY}{dX} = \frac{BH}{EH} = \text{tangent } \theta_s$. The tangent line or hypotenuse (EB) of triangle BHE just touches or skims the parabola at point B .

The problem or job is to find or know where, on the X axis, the point is to which the tangent line from any point on curve is to be drawn. In terms of drawing Fig. 28, how does one know that point E should be just where it is shown to be?

If the trigonometric tangent of the angle of slope is, as above given, equal to $\frac{\sqrt{P}}{\sqrt{X}}$, then, as stated before, the Y (an ordinate) for any point is to be divided by some distance, X (an abscissa), yet to be determined. This yet unknown distance is appropriately to be called the subtangent line (ST). Now the Y or ordinate, at any point of the parabola is $2\sqrt{P}\sqrt{X}$. So the trigonometric tangent of the angle θ in general is

$$\frac{\text{Ordinate}}{\text{Abscissa}} = \frac{Y}{X} = \frac{2\sqrt{P}\sqrt{X}}{ST} = \frac{\sqrt{P}}{\sqrt{X}}$$

$$\text{Therefore } ST = \frac{2\sqrt{P}\sqrt{X}}{\frac{\sqrt{P}}{\sqrt{X}}} = 2X$$

The sub-tangent distances in Figure 28 are DF , EH , and GT . The ST in above equations stands for subtangent in general. This ST must not be taken as the diagonal distance from point S at upper right of figure to the point T on the X axis directly below point C on the parabola at upper right.

Thus, one has, with the aid of the differential calculus and simple trigonometry, arrived at a very important truth: The subtangent is always $2X$ whatever value of X is concerned.

If the subtangent line is, in general, $2X$, then, in a particular case, one goes to the left of the origin, (0) for a distance of X and thereby locates the point to which the tangent line from the point on the curve, whose abscissa is X , is to be drawn. In mirror drawing, the point B has an abscissa X of 16 units. Therefore the point E is at an equal distance of 16 to the left of the origin (0).

Tangent angle

$$\theta_s = \frac{\sqrt{P}}{\sqrt{X}} = \frac{BH}{EH} = \frac{16}{32}$$

$$\theta_s = 26^\circ 34'$$

Again, the tangent line at point A , with an X of 4, is drawn to point D , which is 4 units to the left of the origin, (0). The angle θ_A would be the angle involved. For point C with an X of 36 units, the tangent line at C is drawn to point G which is 36 units to the left of the origin, (0). The angle θ_C is thus involved. Of course θ_A is a greater angle than θ_s which, in turn, is greater than θ_C .

$$\text{Tangent angle } \theta_A = \frac{AF}{DF} = \frac{8}{8} : \theta_A = 45^\circ$$

$$\text{Tangent angle } \theta_C = \frac{CT}{GT} = \frac{24}{72} : \theta_C = 18^\circ 26'$$

The gist and crux of the proof, of the major objective, so far has been provided by the differential calculus. The equation of the parabola is $Y = 2\sqrt{P}\sqrt{X}$. The calculus says that in general the trigonometric tangent $\frac{dY}{dX}$ is $\frac{\sqrt{P}}{\sqrt{X}}$. Thus one is enabled to know that the value of the subtangent is $2X$ and therefore enables one to locate the "western" terminus of the tangent line on the X axis.

There is an important observation to be made at this juncture. From inspection of the parabola, one knows that the slope angle along the curve decreases as one goes out along the curve. The calculus tells us specifically that the trigonometric tangent of the slope angles is inversely proportional to the square root of the abscissa X of the point as the point is out farther along the curve.

THE PERPENDICULAR OR NORMAL LINE AT ANY POINT ON A PARABOLA

The next job at hand is to determine how to draw the perpendicular or normal to the tangent line at any point along the parabola. When we do find out, there will be two right triangles with a common side. This means, in the mirror drawing, as one considers point B , that right-angled triangle EBH is similar to triangle HLB . In turn, this means that angle HBL is equal to angle θ_s . Therefore the trigonometric tangent of θ_s equals the tangent of angle $HBL = \frac{HL}{HB}$. We know, in accordance with the differential calculus, that

$$\tan \theta_s = \frac{\sqrt{P}}{\sqrt{X}} = \frac{HL}{HB}$$

We know also that BH or the ordinate of point B or any point on the parabola is $Y = 2\sqrt{P}\sqrt{X}$. The normal line at point B is line BL . The line HL along X axis is called the subnormal, SN . We want to know where to locate point L in particular or to know the length of the subnormal for any point on parabola in general.

We can now write:

$$\tan \theta_s = \frac{\sqrt{P}}{\sqrt{X}} = \tan \text{angle } HBL = \frac{HL}{HB} = \frac{SN}{Y} = \frac{SN}{2\sqrt{P}\sqrt{X}}$$

Therefore,

$$SN = \frac{2\sqrt{P}\sqrt{X}\sqrt{P}}{\sqrt{X}} = 2P.$$

This simple final expression provides the interesting and important fact that the subnormal for all points on the parabola is the same and is equal to a readily known distance, $2P$. So the subnormals for the particular points A , B and C on curve in mirror drawing are FK , HL and TM , respectively, and are constant in length (8 units). We are now able to draw perpendiculars or normals from points A , B and C to points K , L and M respectively.

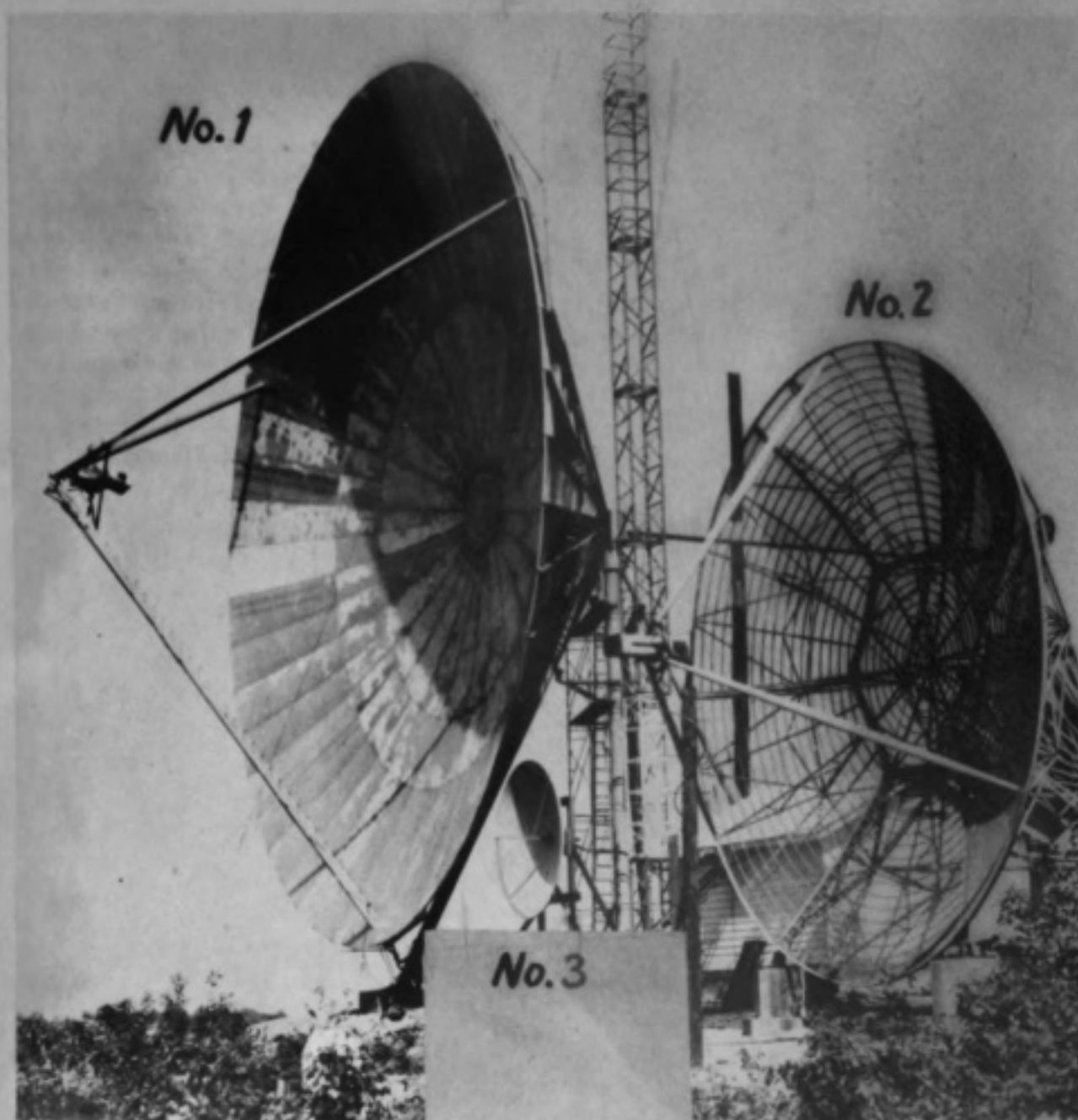
ANGLE OF INCIDENCE EQUALS ANGLE OF REFLECTION AT EVERY POINT ON PARABOLA

Having established and rationalized, with the (splendid) aid of calculus and trigonometry, the above basic facts and truths relative to the intrinsic nature of the parabola, we are ready to embark on the last part of the discussion and study of the parabola. It will be a pleasant and satisfying part of the journey. In terms of mirror drawing, we want to show the important and interesting fact that, when rays of sound waves, light waves or radio waves come in parallel to the principal axis of the concave parabolic mirror, they will converge on reflection to the focus of the parabola. Likewise, if a source of sound, light or radio waves is at the focus, all the diverging outgoing rays will reflect parallel to the principal axis. This characteristic is an intrinsic one embodied in the parabola only.

In the analysis of our present objective about equality of angles of incidence and reflection, consider point C for example. A wave ray SC strikes the surface at point C and reflects to focus F . It does so because of a basic and simple physical law of reflection. It must be, then, that angle SCM is equal to FCM ; that θ_i is equal to θ_r . It is to be established or proved in terms of the intrinsic nature of the parabola as embodied in mirror drawing, that, wherever a wave ray coming in parallel to the principal axis strikes the curve, the ray will reflect to F .

Ray waves from a far distant sound, light or radio transmitter will be parallel; so, if the parabolic reflector is properly aimed, the rays will be precisely focused to a single point.

Since the eccentricity of the parabola is unity (1), the focal distance CF is equal to the directrix distance, CD_c . It is evident that the distance CD_c is the distance TO plus the distance P . The distance TO is equal to X . So focal distance $CF = X + P$. It is also evident that the distance GO to the left of origin is equal to X . (It is to be recalled that the subtangent is $2X$.) In turn, therefore, it is evident that that distance GF is also $X + P$. So $GF = CF$. This means that the triangle CGF is an isosceles triangle. The angle CGF (θ_c) is equal



NO. 1 RIM DIAMETER 60 FT.: FOCAL LENGTH 25 FT.
NO. 2 RIM DIAMETER 28 FT.: FOCAL LENGTH 12 FT.
NO. 3 RIM DIAMETER 8 FT.: FOCAL LENGTH 3 FT.

MICROWAVE RADIO PARABOLOID ANTENNAE
FIGURE 28A

to angle GCF . As per previous statement in connection with point B , that angle θ_s was equal to angle HBL , so it is true that angle θ_c is equal to angle TCM . So, in turn, the angle GCF is equal to angle TCM . Finally, since angle FCM (θ_r) and angle SCM (θ_i) are complementary angles of θ_c in the two right triangles GCM and SCT respectively, these angles are equal. The angle of incidence, SCM (θ_i) is equal to angle of reflection FCM (θ_r).

The argument and analysis appertaining to the situation at point C applies in like fashion to the rays WB and ZA incident at points B and A above the principal axis and to the three rays shown below the principal axis. The reflection at the six points shown in mirror drawing is typical of every point on the parabola when a broad beam of waves from a distant source strikes the surface.

Of course the equal angles of incidence and reflection at a particular point (C , for example) are not equal to the equal angles of incidence and reflection at other points (A and B , for example). The equal angles of incidence and reflection decrease as the point of reflection nears the vertex point O .

Mathematics of several kinds (algebra, arithmetic, trigonometry, plane geometry, analytical geometry, differential calculus) have been the "tools" which, together with a few simple and valid doctrines, have provided a fine proof of the focusing nature of a parabolic surface. In many practical items of sound, light and radio equipment, it is highly desirable to have a mirror which will accurately focus the sound, light and radio waves coming from a distant source. One is confronted with many types of parabolic reflectors on many fronts in one's everyday environment.

As stated above, mathematics of several kinds are excellent and useful "tools" with which to work. The reader can look forward to further exploration of the easy essentials of these tools and their practical utility.

CHAPTER 5

The Hyperbola

In the family of curves called the conic sections, the hyperbola completes the "round up" of such curves from the standpoint of eccentricity. It was pointed out previously that the term eccentricity can be applied to the circle and is zero (0). For the hyperbola, the eccentricity is greater than unity (1).

It is suggested to the reader that he look back at photographs page 14, (Figs. 7 and 8) and note the manner with which a plane intersects a right circular cone to provide a hyperbola.

The hyperbola provides interesting and provocative study. A particular type of this curve, called the equilateral or rectangular hyperbola, serves as the basis for the hyperbolic functions $\sinh$, $\cosh$ and $\tanh$.

Before discussion and exposition of the intrinsic nature, properties and usefulness of these functions, it is in order to make a few general observations regarding the curve itself.

The hyperbola curve lays out the path or route along which an object might move under very definite specifications. If one takes a journey on a hyperbolic path, then one surely does "go into the wide blue yonder". It is an open curve. Even if one went to infinity on a hyperbola he would never get back, which one might do theoretically if his journey were made along a parabola. A hyperbolic journey, like a parabolic journey, cannot be repeated; there can be no periodicity or recurrence as there can be with the circle and ellipse. This latter point is an important one to keep in mind.

FREQUENCY VERSUS WAVE LENGTH IN WAVE COMMUNICATION

It would be interesting to inquire if a hyperbola is ever actually encountered in technological endeavor. So, aside from saying, at this very beginning of the discussion that the eccentricity (e) of a hyperbola is greater than unity (1), an excellent and practical example of the in communication hyperbola will be discussed at once. See accompanying chart Fig. 29.

WAVE LENGTH, FREQUENCY, VELOCITY

Sound waves and electromagnetic waves (guided by wires and radio waves) are the two main types of waves used in communication of information. Light waves are a restricted portion of the electromagnetic spectrum. In both types of waves, a simple relation exists among wave length, frequency and velocity. This relation is nicely expressed in a simple formula

$$\text{Wave Length (L) times Frequency (F) = Velocity (V)}$$

$$LF = V = A \text{ constant}$$

For electromagnetic waves, this equation becomes

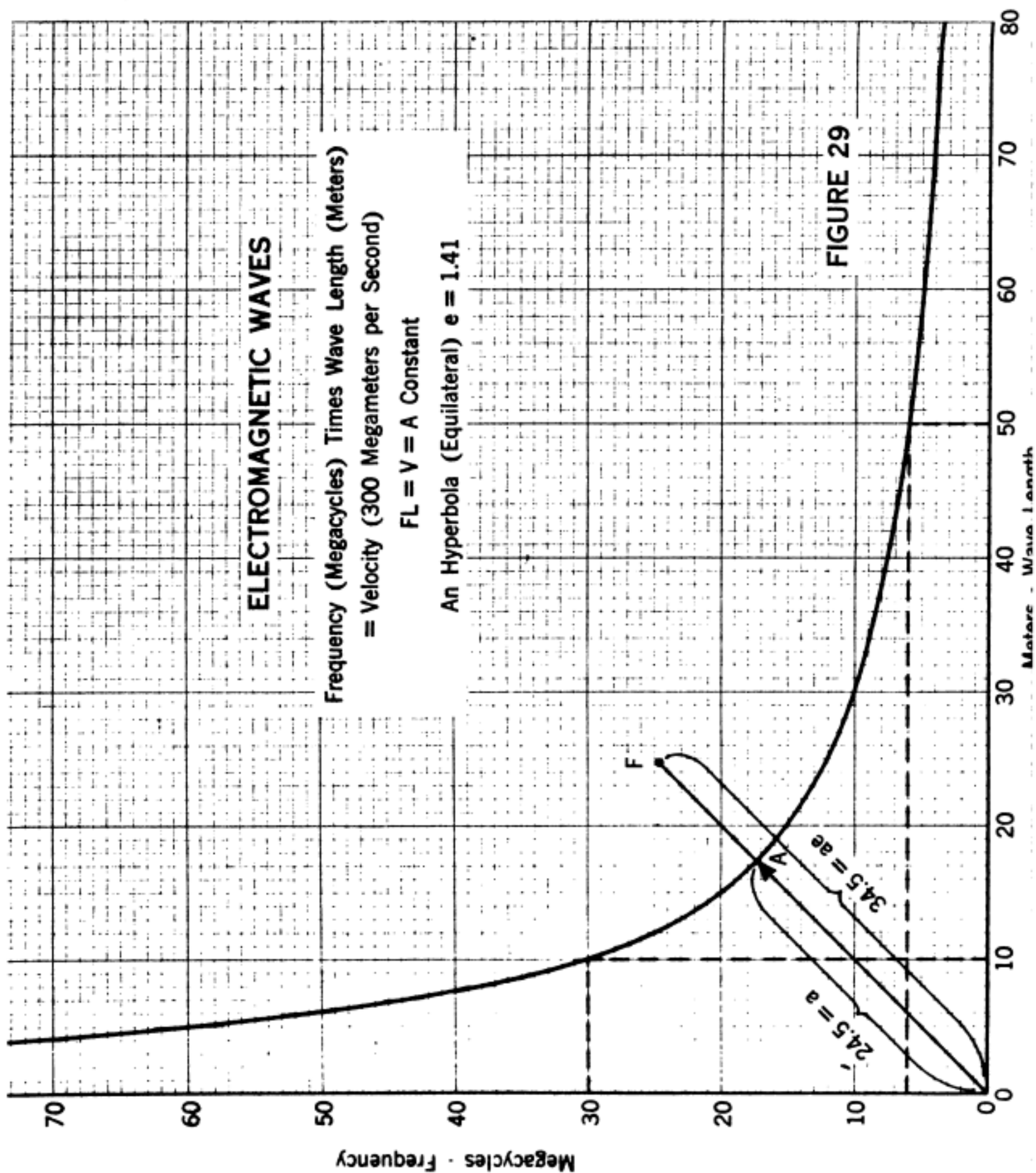
$$L \text{ (meters)} F \text{ (megacycles)} = 300 \text{ megameters per second}$$

300 megameters per second is 186,000 miles per second.

This equation says that the product of wave length and frequency is always the same, a constant. In terms of conventional coordinate axis notation.

$$XY = V = A \text{ constant}$$

If we plot F on Y axis and L on X axis, we will get a curve. This curve is hyperbola; an equilateral and rectangular hyperbola with its asymptotes as the coordinate axes.



It will be seen that if the meters are few (say 1 meter on X axis), the point on curve for the corresponding frequency (300 megacycles) would be way up high. Likewise for 1 megacycle on Y axis, the point on curve for the corresponding wave length (300 meters) would be way out to the right.

The curve at its upper left and its lower right would gradually get closer and closer to the Y axis and X axis respectively and (as the saying is) would touch these asymptotic axes at infinity ("out in the wide blue yonder").

Plotting the curve as has been done, the X and Y axes are actually the asymptotes. In other words, the equation of the hyperbola is ($XY = \text{a constant}$) only when it is referred to its asymptotes as the coordinate axes. Another equation will be discussed later involving other axes.

GRAPHICAL INTERPRETATION; AREAS

Ever since Pythagoras and his famous theorem and his Greek confreres, it has been good practice to regard curves from the point of view of areas. In the hyperbolic curve herewith shown, it is of interest to observe that the two rectangular areas shown with two sides as dashed-lines are equal in area, namely 300. If a rectangle were thus drawn for each and every point on the hyperbola, the areas would always be the same. This is what the equation $XY = V$ says and the equal rectangles portray the idea graphically. On the curve chart, two areas bonded by dotted lines are designated. $10 \text{ times } 30 = 6 \text{ times } 50 = 300$.

Again it is to be observed and emphasized that a hyperbolic journey (like a parabolic journey) is not periodic and cannot be repeated. The hyperbola lays out a course from some starting point to infinity and then it is all over; the traveler never comes back.

MAIN AXIS OF HYPERBOLA; FOCUS

If a diagonal line is drawn from origin (O) at 45 degrees to X and Y coordinate axes of the accompanying curve ($FL = V$), one would have the main axis as it were. The distance from origin (O) to point of intersection with hyperbola curve, called the vertex, (A), is conventionally designated by the letter (a); $OA = a$. The point " A " on curve actually has the coordinates, $\sqrt{300}$; namely 17.3 for X and 17.3 for Y . Therefore " a " or $OA = 24.5 = \sqrt{2} \text{ times } 17.3 = 1.41 \text{ times } 17.3$.

As previously stated and yet to be discussed more completely, the eccentricity (e) of a hyperbola is greater than unity (1). It will be merely stated at this point in the presentation, that in the particular hyperbola above (sometimes called an equilateral or rectangular hyperbola) the eccentricity (e) is actually 1.41 ($\sqrt{2}$).

At this stage of the discussion (later to be argued about more completely, it is to be stated that the distance from the origin of coordinate axes (actually the hyperbolas' asymptotes in this particular case) to the focal point (F) of the curve is the distance ae or $1.41a = 1.41 \cdot 24.5 = 34.5$.

The focus (F) has been so located by actual measurement on the frequency-wave length plotted hyperbola.

GENERAL SPECIFICATION AND DEFINITION FOR HYPERBOLA

Now let us return to hyperbolas in general.

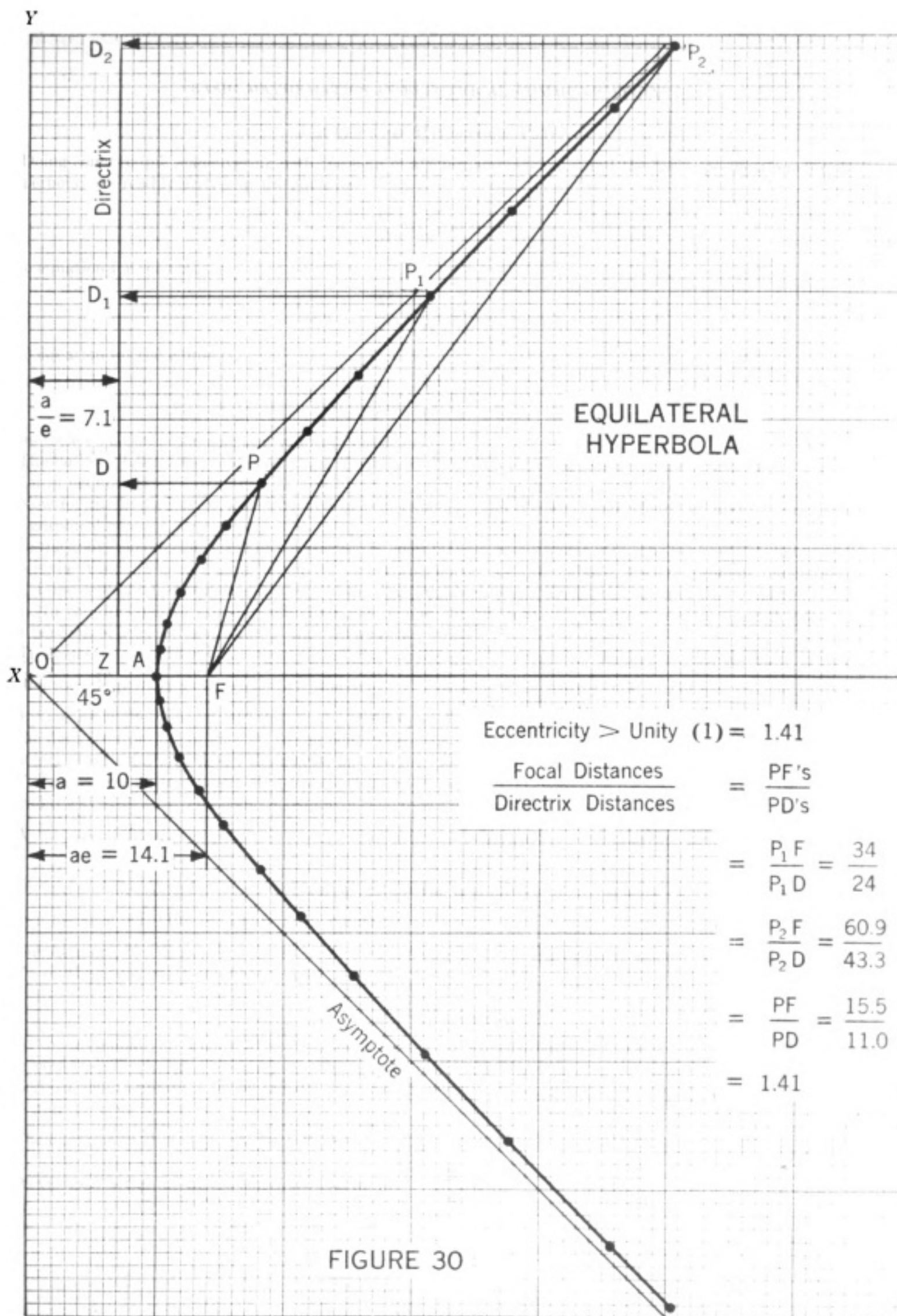
The specification for a hyperbolic path is that wherever you are on the curve, the distance to a fixed point (focus) is always greater than the perpendicular distance to a fixed reference line (directrix). In other words, the eccentricity is $1+$ rather than $1-$, as for the ellipse and 1 as for the parabola.

To write the specification differently, it is to be stated that the ratio of the focal distance to the directrix distance is the same sort of ratio (e) as for ellipse and parabola, but, in the case of the hyperbola, the ratio (e) is greater than unity (1).

In terms of the accompanying Figure 30 the specifications say that PF is greater than PD and PF is greater than P_1D_1 .

$$\frac{PF}{PD} = \frac{P_1F}{P_1D_1} = \text{a constant} = e = \text{eccentricity} = 1.41$$

As stated, the eccentricity for the equilateral or rectangular hyperbola is the $\sqrt{2}$ or 1.41; the accompanying sketch is such a hyperbola.



EQUATION OF HYPERBOLIC PATH

Doing some simple maneuvering with the help of algebra and geometry and the basic specifications for the hyperbola, one readily arrives at the equation

$$\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1.$$

The equation for the circle is

$$\frac{X^2}{R^2} + \frac{Y^2}{R^2} = 1.$$

The equation for the ellipse is

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1.$$

R in circle equation is the constant radius of circle.

The " a " in ellipse is the semi-axis along X axis and the " b " is the semi-axis along the Y axis. They are not equal in magnitude.

Now for the hyperbola, it would be well to very explicitly find out what and where " a " and " b " are, from the graphical point of view.

In the above sketch, the origin or intersection of coordinate axes is at point " O ". The horizontal or "abscissa" axis is the X axis. The vertical or "ordinate" axis is the Y axis. As stated above, it would promote understanding, if one could find some actual distance in the drawing which is graphically the " a " and the " b " of the equation. In the drawings of a circle and ellipse, the R and the " a " and " b " are easily identified graphically. In the drawing of a hyperbola, where is the " a " and the " b "?

To geometrically and graphically find out what distance is " a " and what distance is " b ", the basic specification for a hyperbola as shown in first sketch $\frac{PF}{PO} = e$, where " e " is greater than 1, will provide the answer.

By means of the basic specification and some simple algebra, one readily comes up with the fact that " a " is the distance from the origin " O " to the point where the curve crosses the X coordinate axis. The " a " therefore is a simply and graphically designated and definite distance. The " a " is marked on sketch; it is OA . This " a " distance was given numerically for the equilateral or rectangular hyperbola for wave length and frequency of radio waves as 24.5. The point A , where curve crosses coordinate X axis, is called the vertex of the curve.

As per specifications for point " A " which is of course a point on the accompanying general curve, $\frac{AF}{AZ} = e = \text{eccentricity}$.

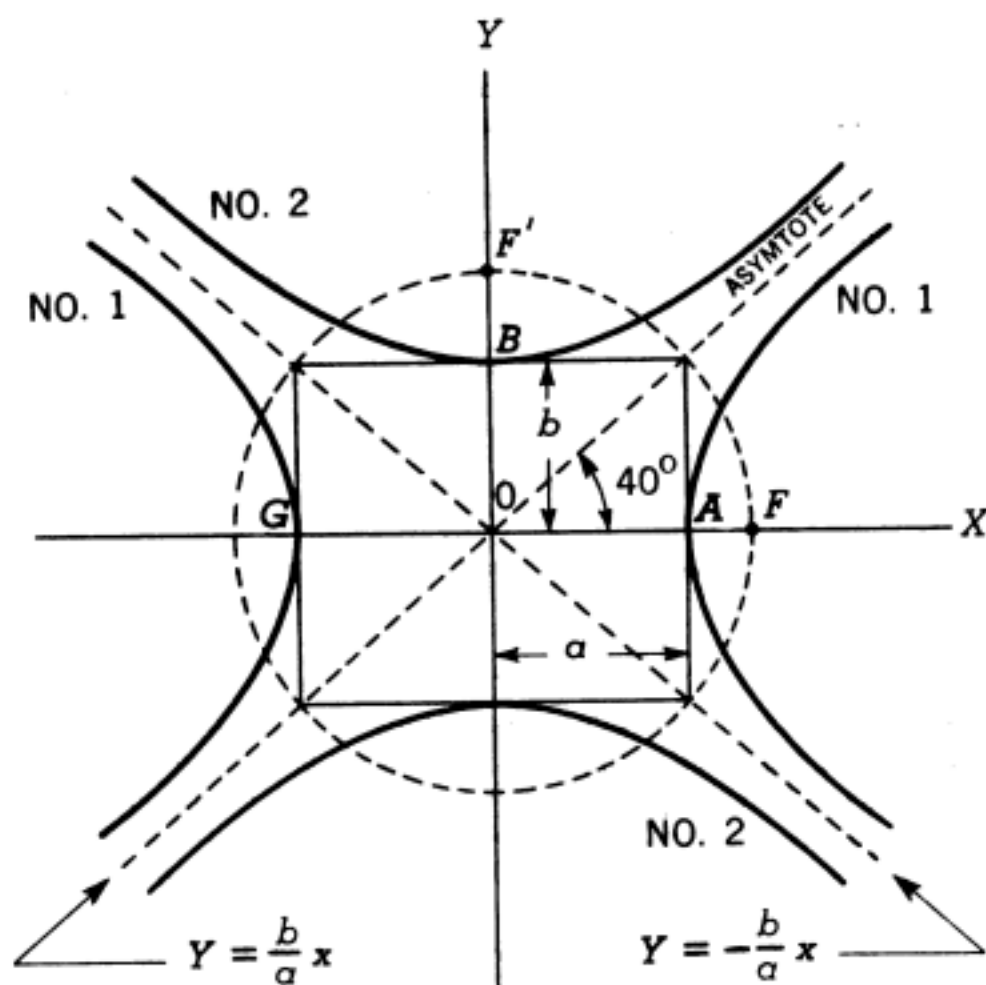
In terms of the X and Y axes, which now are not the asymptotes of this general curve, the directrix line is at a distance of $\left(\frac{a}{e}\right)$ to the right of the Y axis.

Furthermore, the distance of the focus F from the origin " O " is the distance ae , and is 1.41 times 24.5 or 34.54 on the wave length-frequency curve.

CONJUGATE HYPERBOLAS

In the accompanying Figure 31 sketch, the X axis is the so called principal axis of the hyperbola at hand. One could readily imagine a companion hyperbola drawn with the Y axis as the principal axis. This latter hyperbola could appropriately be called a conjugate hyperbola. The vertex of this conjugate curve is at point " B " and its focus is at (F').

The equation of this conjugate hyperbola is $\frac{Y^2}{b^2} - \frac{X^2}{a^2} = 1$. For this curve, the distance from the origin " O " to its vertex B is the " b " we have been looking for in $\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1$.



CONJUGATE HYPERBOLAS

"a" is greater than "b"

$$OF = OF'$$

$$\text{NO. 1} \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

$$e = \frac{\sqrt{a^2 + b^2}}{b} = 1.31$$

$$\text{NO. 2} \quad \frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$$

$$e = \frac{\sqrt{a^2 + b^2}}{a} = 1.53$$

FIGURE 31

So for the conjugate hyperbola, the distance BO or " b " serves in its own equation and gets or borrows as it were its " a " or AO from its companion or conjugate. The converse statement readily follows for the $\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1$ hyperbola with which the discussion started.

One final good point to be made is the reason for calling the two hyperbola conjugate. If the distance from O to F , origin to focus, is used as a radius (see curves) of a circle and a compass swings around about (O) from OF position toward Y axis, it will cut Y axis at a certain point. This point will be the focus (F') for the vertical axis curve. This vertical axis curve is the conjugate and the same basic specification for hyperbolas in general apply to both hyperbolas. They do have the same focal distances $OF' = OF$ but not the same equation; the a 's and b 's are interchanged.

Particularly it is to be mentioned that the conjugate hyperbolas in general do not have the same magnitude of eccentricity. It will be seen in curves shown that the vertical hyperbola spreads out more. It comes about rather readily that the eccentricity (always greater than 1 in every case) for X axis hyperbola is

$$e = \frac{\sqrt{a^2 + b^2}}{a}$$

The eccentricity for the conjugate Y axis hyperbola is

$$e = \frac{\sqrt{a^2 + b^2}}{b}$$

Since in the particular curves shown, " a " is longer than " b ", the eccentricity (e) for X axis hyperbola is smaller than the (e) of its conjugate hyperbola on Y axis.

ASYMPTOTES

The dotted lines in the accompanying Fig. 31 are the asymptotes. They are the lines which at a great distance (infinity) out on the hyperbolas would be tangent to the curve; just touch or meet the hyperbola at a very, very small angle. Even though the hyperbolas are conjugate, they have different e 's it will be recalled. Also it will be noted by inspection that the asymptotes of the two hyperbolas are not at right angles to one another. The angles formed by asymptotes and X axis is 40° .

EQUILATERAL OR RECTANGULAR HYPERBOLAS

If two hyperbolas are laid out on X and Y axes by intention so that $a = b$ then one gets conjugate hyperbolas which are called equilateral or rectangular. They do have the same eccentricity (1.41) and the same shape. Their asymptotes would be right angles and would make a 45° angle with the X and Y axes. They have same focal distance (OF) and particularly again they have the same distance from vertex to (O); $a = b$ and $b = a$.

REVIEW OF FUNDAMENTAL EQUATION OF HYPERBOLA

As previously stated, the basic equation for the hyperbola is

$$\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1 = \left[\left(\frac{X}{a} \right)^2 - \left(\frac{Y}{b} \right)^2 \right]$$

The " a " is the distance along X axis from vertex to " O " the origin of coordinate axes. The " b " is the distance along Y axis from vertex of the conjugate hyperbola to " O " the origin of coordinate axes. For every point on hyperbola curve, from vertex " A " out northeastwardly as it were, a comparison or ratio of its X distance, (the abscissa) is made to the " a ", (an abscissa too) measured also along the same X axis, namely $\frac{X}{a}$. For the same point on curve, a comparison or ratio of its Y distance (the ordinate) is made to the " b " (an

ordinate too) namely $\frac{X}{b}$. In this case, the "b" or ordinate is borrowed as it were from the conjugate hyperbola on Y axis. Then one wrote the equation involving the squares of these two ratios; $\left(\frac{X}{a}\right)^2 - \left(\frac{Y}{b}\right)^2 = 1$.

It is in order to refer back to ellipse and note that the same kind of ratios were considered; namely squares of ratios between two abscissae, $\left(\frac{X}{a}\right)^2$ and two ordinates $\left(\frac{Y}{b}\right)^2$ to get above equation.

However, it will be remembered that for the circle, the ratio of X to the hypotenuse of the circle (not an abscissa) was taken. Likewise the ratio of Y (an ordinate) to the hypotenuse of the circle (not an ordinate) was taken. The hypotenuse was a single constant distance which offered its services for comparison for all the angles.

The simple equation is $\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1$.

Then $\frac{X}{R}$ was called a circular cosine for an angle and $\frac{Y}{R}$ was called a circular sine of an angle. Finally then we wrote $\cos^2 \theta + \sin^2 \theta = 1$.

EQUILATERAL OR RECTANGULAR HYPERBOLA

If as said before " a " = " b ", then the particular hyperbola (two conjugate hyperbolas really) becomes an equilateral or rectangular hyperbola, see Figure 32.

Its equations would be

$$\frac{X^2}{a^2} - \frac{Y^2}{a^2} = 1 = \frac{Y^2}{b^2} - \frac{X^2}{b^2}$$

We could write for the one (X axis) curve,

$$X^2 - Y^2 = a^2 = b^2$$

We could write for other (Y axis) curve

$$Y^2 - X^2 = b^2 = a^2$$

Again, " a " is equal in numerical value to " b "

If " a " and " b " are assumed to be unity (1) then

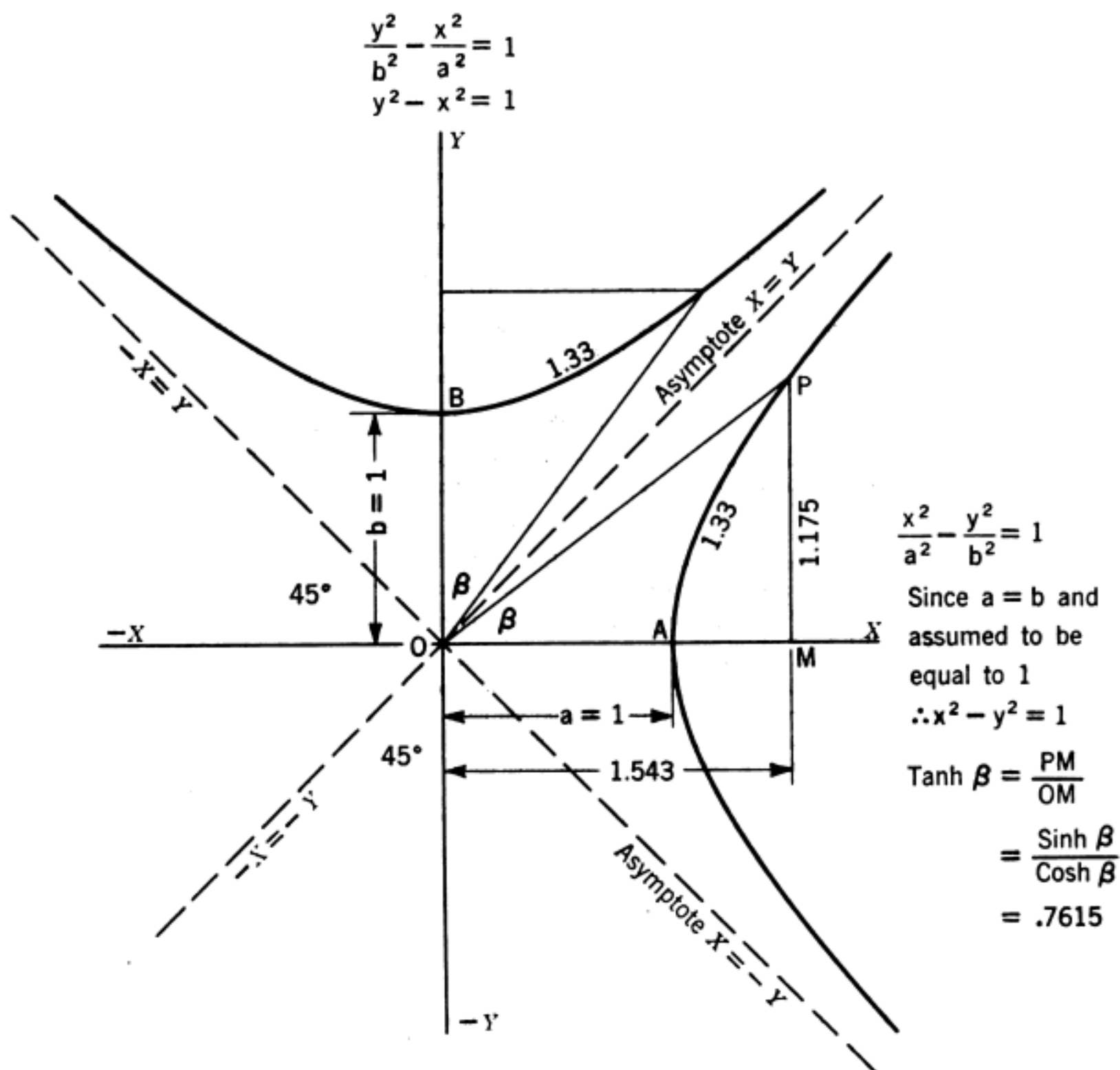
$$X^2 - Y^2 = 1$$

$$Y^2 - X^2 = 1$$

These are the basic and simple equations of two conjugate equilateral or rectangular hyperbolas.

Unlike the theorem of Pythagoras, these equations state a difference in two squares.

CONJUGATE EQUILATERAL HYPERBOLAS



$$\frac{PM}{b} = \frac{PM}{a} = \frac{y}{a} = \sinh \beta = \frac{1.175}{1} = 1.175$$

$$\frac{OM}{a} = \frac{x}{a} = \cosh \beta = \frac{1.543}{1} = 1.543$$

Hyperbolic Angle $\beta = 1$ Radian

FIGURE 32

CHAPTER 6

Hyperbolic Sines and Cosines

All of the preceding discussion in this treatise and the (it is hoped) simple logical arguments and graphical interpretations about circles, ellipses, parabolas and hyperbolas have been aimed at understanding and getting the "feel" of hyperbolic sines and cosines and an insight into their practical use in electrical engineering.

The specific and particular basis for defining these hyperbolic sines and cosines is furnished by the simple geometry of the equilateral or rectangular conjugate hyperbolas and their equations.

$$X^2 - Y^2 = 1 = Y^2 - X^2.$$

The two sets of asymptotes are right angles to one another and at 45° to the coordinate axes.

The doctrine pursued will be much like that followed in the case of circular sines and circular cosines. There will be one important and significant difference.

SINH AND COSH DEFINITION

The ratio of the "X" distance (an abscissa, OM) of a point "P" on a hyperbola to the distance "a" (graphically shown to be distance OA from vertex to origin "O") also an abscissa is called the hyperbolic cosine of the hyperbolic angle, POM . $\text{Cosh } \beta = \frac{X}{a} = \frac{OM}{OA}$, see Figure 32.

The ratio of the "Y" distance (an ordinate, PM) of the point "P" on the curve to the distance "b" (graphically shown to be distance from vertex of conjugate equilateral hyperbola to origin "O") also an ordinate distance, OB , is called the hyperbolic sine of the hyperbolic angle, POM . $\text{Sinh } \beta = \frac{Y}{b} = \frac{MP}{OB} = \frac{MP}{OA}$.

For the hyperbolic cosine, two abscissas are compared and for the hyperbolic sine, two ordinates are compared. The basis of comparison is a constant distance "a" or "b" which are equal in length but "a" is an abscissa and "b" is an ordinate. They are in fixed position as it were and do not rotate as the constant length hypotenuse did in the case of the circle. But again, "a" and "b" are equal in magnitude and are available for comparison to two quantities which are changing, namely the "X" and the "Y", as the point pursues its course out along the hyperbolic path.

$$\text{Sinh } \beta = \frac{Y}{b} = \frac{Y}{a}$$

$$\text{Cosh } \beta = \frac{X}{a}$$

$$\text{Tanh } \beta = \frac{Y}{X} = \frac{MP}{MO}$$

This definition of sine and cosine, really hyperbolic sine and cosine, is a new and different kind of definition in general than that which is used in circles. Two ordinates are compared to get the hyperbolic sine and two abscissas are compared to get the hyperbolic cosine. An hypotenuse plays no part because there is no constant hypotenuse existent. There is a constant distance available to which comparison can be made and that distance is "a" (an abscissa) and an equal distance "b" (an ordinate).

The new or at least different basis for definition is a perfectly valid and tenable one.

So one can write in a similar but strikingly different manner not

$$\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1 = \cos^2 \theta + \sin^2 \theta$$

as is the case for circles where R is an hypotenuse but

$$\left(\frac{X}{a}\right)^2 - \left(\frac{Y}{b}\right)^2 = 1 = \cosh^2 \beta - \sinh^2 \beta \text{ for the hyperbola}$$

Remember that " a " = " b ", but " a " is an abscissa and " b " is an ordinate. As will be observed, hyperbolic sines and cosines are distinguished from circular sines and cosines by adding the letter " h ", to become $\sinh$ and $\cosh$.

HYPERBOLIC TANGENT

Since the circular tangent is ratio of Y distance (ordinate of point " P ") to X distance (abscissa of point " P "), the constant length radius (R) is not involved. In other words, the tangent of a circular angle is ratio of side "opposite" the angle to the side adjacent to the angle. In circular matters

$$\frac{\sin \theta}{\cos \theta} = \tan \theta = \frac{Y}{X} = \frac{\text{an ordinate}}{\text{an abscissa}}$$

So likewise in hyperbolic matters

$$\frac{\sinh \beta}{\cosh \beta} = \tanh \beta = \frac{Y}{X} = \frac{\text{an ordinate}}{\text{an abscissa}}$$

CIRCULAR RADIANs AND AREAS

A circular radian is an angle obtained by the simple doctrine of laying off on the circumference of the circle, a distance or an arc equal numerically to the radius. One says an arc of R subtends one radian of angle in a circle. Since the radius could be laid off on the circle 2π times, then 2π radians is the entire angular sweep around the center. Using the conventional and purely arbitrary system of degrees, there are 360 degrees in a circle; 1 circular radian = $57^\circ 17' 45'' = 57.296^\circ$.

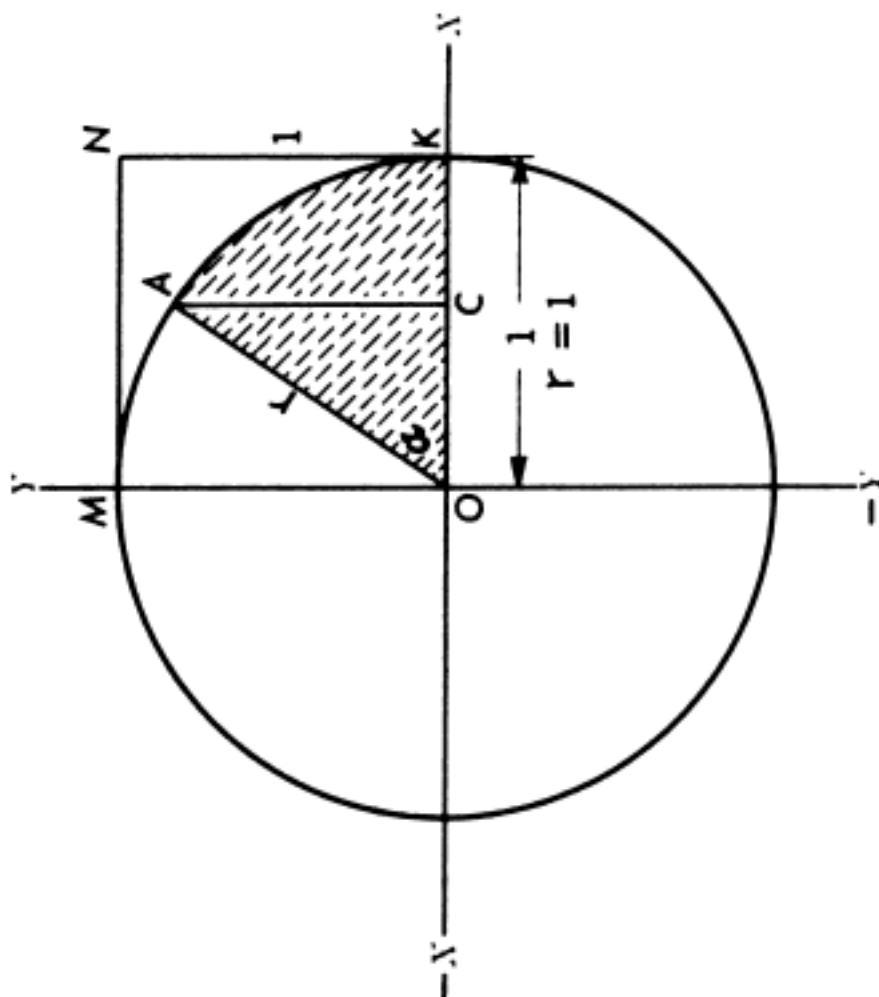
$$\pi = 3.141592$$

Emulating the Greek's inclination to consider geometrical problems from the standpoint of areas, the area of the shaded area in the accompanying sketch Fig. 33 is found to be .5 square inch. This follows simply from the fact that if R (radius) is assumed to be unity (1), then the area of the circle (πR^2) is π square inches. The total π square inches divided in 2π sectorial areas (like small pieces of apple pie) means that each sector is $\frac{1}{2}$ or .5 square inch in area. In general, each circular radian area is $\frac{R^2}{2}$. If $R = 1$, area = $\frac{1}{2}$. If $R = 10$, area = 50.

In the sketch of the circle, Figure 33 the angle α is a circular radian and its sine, $\frac{AC}{AO}$ is .841 in value. Since OA (R) is assumed to be unity (1), AC must be .841 inch and numerically equals the sine. The cosine of one radian, $\frac{OC}{OA}$ is .540. Again, since OA (R) is assumed to be unity, the cosine of one circular radian is OC , or .540. In the sketch, the distances are not actual inches if one would measure them by a ruler, but the arguments are the important factor and the actual size and scale of the drawing makes no difference.

The principal point of the above argument is that a "piece of pie shaped" sector of a circle of .5 square inch in area defines a second time, as it were, or sets up the specification for one circular radian angle. This idea was brought about by originally defining a radian as that angle which is subtended by an arc of R (radius) in length. ($R = \text{arc } AK$.) R was assumed to be unity (1) for the sake of simplicity. One could say that using the arc as a basis for definition is secondary but the .5 square inch area basis for definition is primary. The latter takes precedence over the former. Area $OAKO = \frac{1}{2}$ area $OMNKO$.

CIRCLE CIRCULAR RADIANT AND AREA



$$\begin{aligned}\tan \alpha &= \frac{AC}{OC} \\ &= \frac{\sin \alpha}{\cos \alpha} \\ &= 1.5574\end{aligned}$$

$$\frac{AK^2}{2} = \text{Circular Sectorial Shaded Area} = \frac{r^2}{2} = \frac{1}{2} = \text{OAKO} = \frac{1}{2} \text{ Square Area (OMNKO} = 1)$$

$$\alpha = 1 \text{ Circular Radian} = 57.3^\circ$$

$$\text{Circular Arc (AK)} = r = 1$$

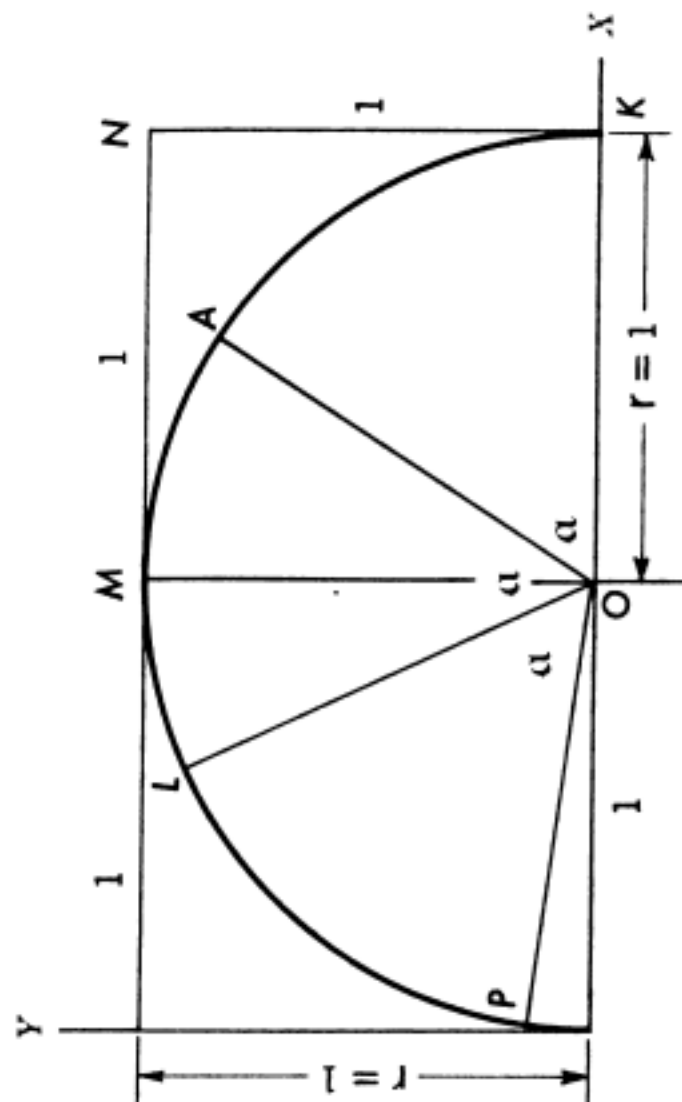
$$AC = \sin \alpha = 0.84147$$

$$OC = \cos \alpha = 0.54030$$

$$\cos^2 \alpha + \sin^2 \alpha = 0.54030^2 + 0.84147^2 = r^2 = 1$$

FIGURE 33

MULTIPLE CIRCULAR RADIAN AREAS



1 CIRCULAR
RADIAN = $57.3^\circ = \alpha$
2 π RADIAN = 360°

Circular Arcs KA , AL and $LP = r = 1$

$$1 \text{ Circular Radian Area (OKAO)} = \frac{r^2}{2} = \frac{1}{2} \cdot \text{Square Area (OKNMO)}$$

$$1 \text{ Circular Radian Area (OALO)} = \frac{r^2}{2} = \frac{1}{2} \cdot \text{Square Area (OKNMO)}$$

$$1 \text{ Circular Radian Area (OLPO)} = \frac{r^2}{2} = \frac{1}{2} \cdot \text{Square Area (OKNMO)}$$

$$2 \text{ Circular Radian Area (OKALO)} = r^2 = \text{Square Area (OKNMO)}$$

$$3 \text{ Circular Radian Area (OKAMLPO)} = \frac{3}{2} r^2 = \frac{3}{2} \cdot \text{Square Area (OKNMO)}$$

$$6.28 \text{ Circular Radian Area} = 2\pi \cdot \frac{r^2}{2} = \pi r^2 = \pi \cdot \text{Square Area (OKNMO)}$$

= Area of Circle

FIGURE 34A

The second circular sketch Figure 34 shows the "piece of pie shaped" sector divided up in ten equal areas, with arc of each area equal to $\frac{1}{10}R$. Each small angle at the center is .1 circular radian.

HYPERBOLIC RADIANES AND AREAS

In the case of a hyperbola, there is neither a 2π nor a radius rotating counterclockwise to create an array of 2π radians of angle and an area of πR^2 , as there was in the case of a circle.

The hyperbola is a particular kind of path or route along which a point moves farther and farther from a starting point never to return. An infinitely large number of hyperbolic radians and a very great area may be involved.

To find out about hyperbolic radians, one pursues the same line of reasoning as used for circles but approaches the problem solely from the area point of view and does not consider hyperbolic arc for a definitive specification.

WHAT IS ONE HYPERBOLIC RADIAN?

The question undoubtedly arising in one's mind is, how was the point *A* of the curve, of accompanying Figure 35 chart selected so that β was one hyperbolic radian?

The answer is as follows. The point *A* was selected out along the curve far enough so that the shaded area *OAKO* (odd in shape) would be .5 square inches. This doctrine of area as a definitive procedure follows the doctrine established for the circle in defining one circular radian.

One might have picked a point along parabolic curve so that the curved distance *AK* (a sort of arc) would be equal to "*a*" assumed to be unity (1). This basis or doctrine of selection would be similar to picking an arc along the circumference equal to radius *R* which was assumed to be unity (1), in order to get one circular radian. This same doctrine is not valid now and cannot be used as a definitive specification for a hyperbolic radian.

If the selection of location of "*A*" on hyperbola had been made "*a* 1a" arc doctrine, the hyperbolic sector area (shaded area) would not have been .5 square inches; it would have less than .5 square inches as seen by inspection of sketch since arc *AK* is a bit greater than "*a*". In other words, the area would have been less than .5 square inches if arc or distance along hyperbola had been selected as "*a*". Again, arc *AK* is a bit greater than "*a*". By electing to pick "*A*" on the .5 square inch basis for the area or *OAKO*, a hyperbolic radian would be defined and would be arrived at in strict compliance with the doctrine of areas as also used to get a circular radian. Therefore, by this doctrine, the angle β is defined as one hyperbolic radian.

As per sketch, this first hradian happens to be $37^\circ 15'$ circular degrees or .65 circular radians. The value of a hyperbolic angle in circular degrees is only interesting and has no particular use or importance. Subsequent consecutive hradians, if one numbers them (1, 2, 3, 4, etc.), get smaller and smaller in circular degrees equivalent; but all subsequent hyperbolic radial areas are .5 square units. The areas get longer and narrower. If "*a*" = 1, area = $\frac{1}{2}$. If *a* = 10, area = 50. Incidentally, the arcs involved get longer and very much longer. Take a look at the chart next to page 77, Figure 37.

The chart Figure 36 shows one hyperbolic radian area divided into ten equal parts. The individual arcs for each .1 hradian increase a wee bit, which is unlike the situation for individual arcs of the .1 circular radians which are equal. The hyperbolic sectors corresponding to each .1 hradian are equal in area (0.5 square inches).

Each .1 hradian area is .1 of the 1 hradian irregular shaped area, *OAKO*, which in turn is .5 of the square area *OMNKO*. Therefore, each .1 hradian area is $\frac{1}{10}$ (.05) of the square area. It is interesting to observe that the first or bottom .1 hradian area is practically a right-angled triangle. Its height is .1 and its base is 1, so that its area is actually $\frac{1}{20}$ (.05) of *OMNKO*. One can imagine 20 of these little triangles comprising the square area *OMNKO*.

It is to be emphasized that the .1 hradian areas are all equal and each is .05 ($\frac{1}{20}$) of the square area *OMNKO* and each is .1 of the area *OAKO*.

NUMERICAL VALUE OF SINH AND COSH FOR ONE HYPERBOLIC RADIAN

Previous sketches (pages 63, 68) portray an equilateral or rectangular hyperbola. It is to be emphasized that the sinh and cosh functions are based specifically on this particular kind of hyperbola. Its equation is

EQUILATERAL HYPERBOLA HYPERBOLIC RADIAN AND AREA

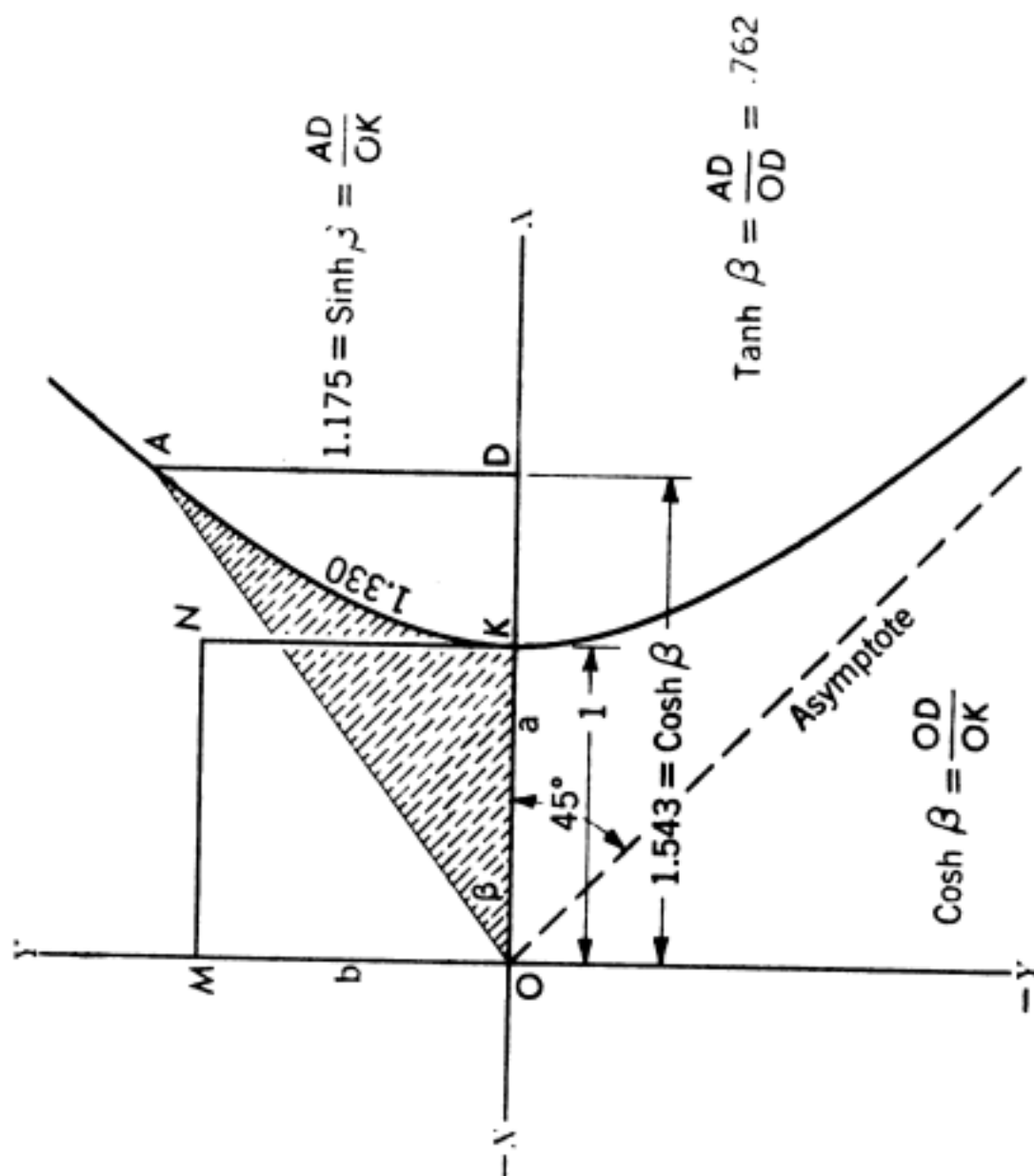
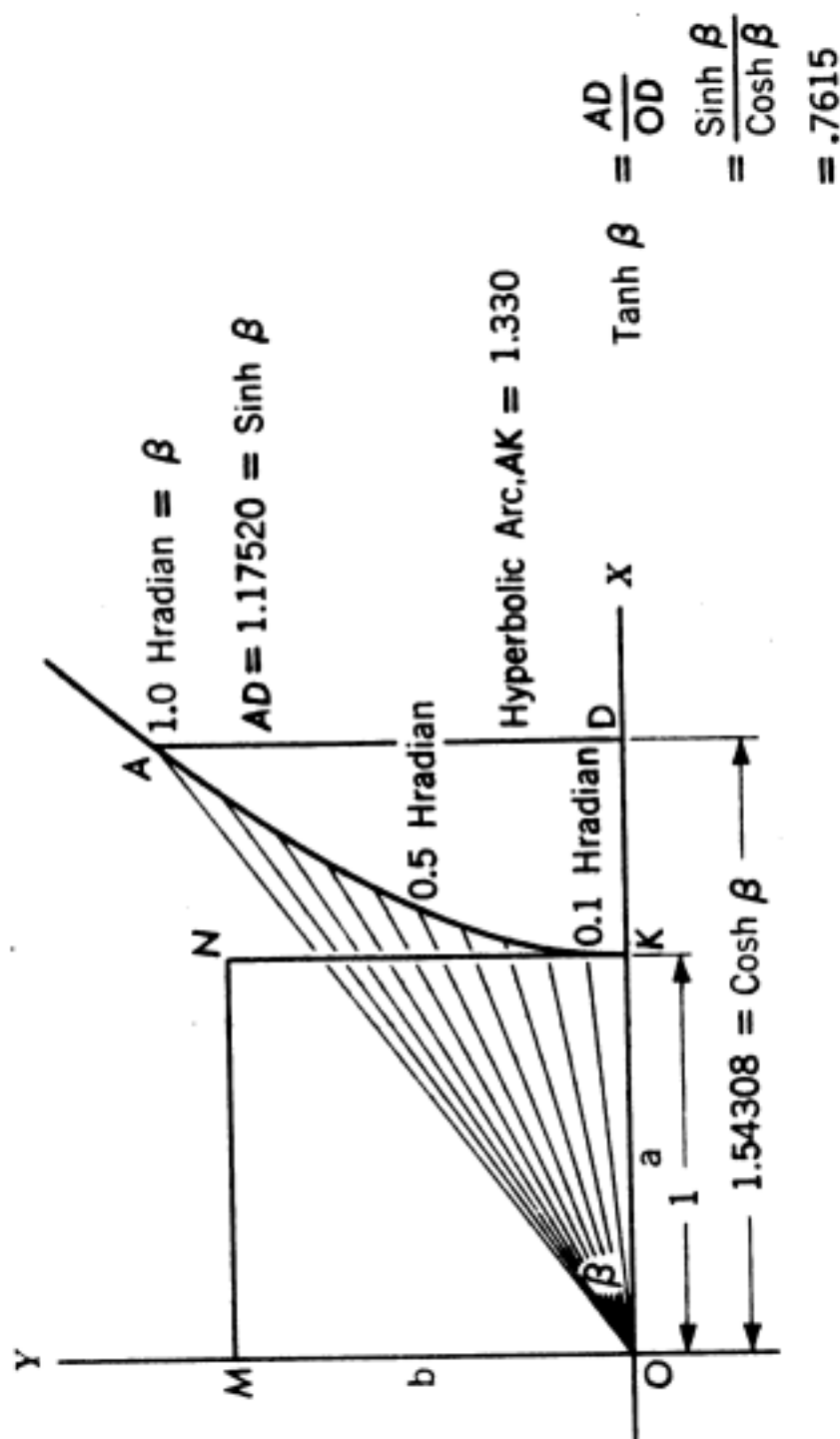


FIGURE 35

EQUILATERAL HYPERBOLA HYPERBOLIC RADIAN



1 Hyperbolic Radian Area = $\text{OAKO} = \frac{1}{2}$ Square Area ($\text{OMNKO} = 1$)

.1 Hyperbolic Radian Area = $\frac{1}{10} \times \frac{1}{2} = \frac{1}{20}$ Square Area

$\beta = 1$ Hyperbolic Radian

$$\text{Cosh}^2 \beta - \text{Sinh}^2 \beta = 1.54308^2 - 1.17520^2 = a^2 = 1^2 = 1$$

FIGURE 36

$X^2 - y^2 = 1$, which is the form of the equation $\left(\frac{X}{a}\right)^2 - \left(\frac{Y}{b}\right)^2 = 1$, resulting if $a = b$ and if both a and b are assumed to be unity (1).

SINH β

The sinh of the hyperbolic angle β is $\frac{AD}{MO}$ or $\frac{AD}{b}$ which is the ratio of ordinate $AD(Y)$ to another ordinate of constant value (b). It is to be noted again that the ordinate " b " is really taken from the conjugate hyperbola; one might say, borrowed from the conjugate hyperbola on Y axis.

Since " b " is MO and assumed to be unity (1), AD is equal 1.175 in length. AD or Y is $\sinh \beta$ graphically. The sinh is 1.175. The angle β is one hyperbolic radian.

COSH β

The cosh of the hyperbolic angle, β , is $\frac{OD}{OK}$ or $\frac{OD}{a}$ and is the ratio of an abscissa (OD or X) to another abscissa of constant value, " a ".

Since " a " = " b " = OM is OK and assumed to be unity (1), OD is equal to 1.543 in length. OD or X is $\cosh \beta$ graphically. The cosh of β is 1.543. The angle β in this case is one hyperbolic radian. The hyperbolic angle is indicated by β ; the beta of the Greek alphabet. The numerical values of $\sinh 1$ as 1.175 and of $\cosh 1$ as 1.543 cannot be precisely determined by divider and ruler. The beautiful mathematical method by which the values are accurately determined is discussed later.

MULTIPLE HYPERBOLIC RADIANs

The point on circle went round and round and a periodic, recurrent and repeated phenomenon was portrayed thereby. However, in the hyperbola, the route goes off to the northeast ever farther and farther and never turns around to come back home. See Figure 37, page 78.

If one goes out to some point whose Y distance (ordinate) is 3.63 and whose X distance (abscissa) is 3.76, one would have a total area of one square inch or one square unit and the angle would be two hradians. For 3 hradians, ordinate distance (Y) of point on path would be 10.02 and abscissa distance (X) would be 10.07. Total area of hyperbolic sector involved would be 1.5 square inches. For 5 hradians, ordinate distance (Y) of point on path would be 74.20 and abscissa distance (X) would be 74.21. The total sectorial area involved would be 2.5 square inches.

.1 hradian	:	$\sinh \beta$	.100	:	$\cosh \beta$	1.005
1 hradian	:	$\sinh \beta$	1.175	:	$\cosh \beta$	1.543
2 hradian	:	$\sinh \beta$	3.627	:	$\cosh \beta$	3.762
5 hradians	:	$\sinh \beta$	74.203	:	$\cosh \beta$	74.210
10 hradians	:	$\sinh \beta$	11,013.	:	$\cosh \beta$	11,013.

The point involved on hyperbolic path to provide 10 hradians would be way "out in wide blue yonder." The total area would be 5 square units.

See later curves page 85, Figure 38 of hradians versus $\sinh \beta$, $\cosh \beta$ and $\tanh \beta = \frac{\sinh \beta}{\cosh \beta}$ which portray how their values vary.

NUMERICAL VALUES OF COSH AND SINH

It is to be observed that even at a rather small number of hyperbolic radians, of which there are an infinite number, the numerical value of cosh and sinh are nearly equal. The reason for this is apparent when one considers their intrinsic nature. Inspection of the curves on accompanying chart Page 115 reveal how the values of cosh and sinh vary as a function of hradians.

Actually, of course, the value of $\cosh$ and $\sinh$ can never be exactly equal since $\cosh^2$ minus $\sinh^2$ must ever be equal to unity (1). For example, the $\cosh^2$ or one hradian is $(1.54308)^2$ or 2.381106. The $\sinh^2$ for one hradian is $(1.17520)^2$ or 1.381395. The difference is very close to unity (1). For six (6) hradians, the $\cosh^2$ is $(201.7156)^2$ or 40,689.18328 and the $\sinh^2$ is $(201.7132)^2$ or 40,688.15054. The difference is very close to unity (1).

The $\cosh$ for 10 hradians is 11,013.232919 and the $\sinh$ is 11,013.232875. Even though these two numbers are very nearly equal, there must be a difference of unity for $\cosh^2 - \sinh^2$.

HYPERBOLIC AREAS AND HYPERBOLIC RADIAN

The area of the three sided region bounded by the radius vector or straight line from origin (O) to a point P on the northeast part of an hyperbola, the " a " on X axis, and the hyperbolic curve itself (See Figures 35, 36, 37) is given by the following formula.

$$\text{Area} = A = \frac{ab}{2} \log_e \left(\frac{X}{a} + \frac{Y}{b} \right)$$

Again with his fine skill, the mathematician makes an "assist" and provides a useful "tool" in his derivation of this formula.

In an equilateral or rectangular hyperbola, the particular curve on which $\sinh$ and $\cosh$ are based and predicated, the " a " equals " b ". If we assume for sake of simplicity that " a " and " b " are equal to one inch, then

$$\text{Area above described} = \frac{1}{2} \log_e (X + Y)$$

The X and the Y are the coordinates of the point A and are numerically equal to the $\sinh$ and $\cosh$ of the hyperbolic angle (B) involved.

If we wish to find the area involved when the hyperbolic angle is one (1) radian, we look in the table for $\sinh$ or Y or one hyperbolic radian (1.1752) and the $\cosh$ or X (1.5431)

$$\begin{aligned} \text{Area} &= \frac{1}{2} \log_e (1.1752 + 1.5431) \\ &= \frac{1}{2} \log_e (2.7183) \end{aligned}$$

The subscript " e " in this equation is the base of the Napierian system of logarithms, 2.7183.

It is at once observed that the 2.7183 (the sum of X and Y) is " e " itself, the base of the Napierian logarithms above mentioned. Of course the $\log_e e$ is unity (1); the log of e to the base e is unity (1). The area in question is therefore one half square inch

$$\text{Area} = \frac{1}{2} \log_e e = \frac{1}{2} \text{ of } 1 = .5$$

It will be remembered that a hyperbolic radian was defined as the angle which would provide a .5 square inch sector. (See Figs. 35, 36). We have had here a good check of the basic doctrine and definition.

If we chose a point, P , farther out on the curve so that two hradians were involved we could readily find resulting area.

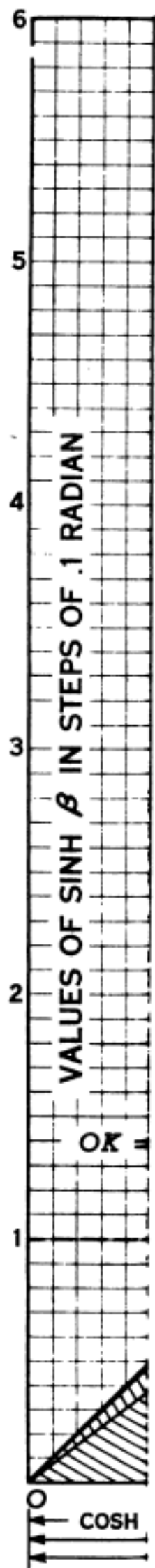
$A = \frac{1}{2} \log_e (X + Y)$, where Y is the $\sinh$ of two hradians and X is the $\cosh$ of two hradians.

$$\begin{aligned} A &= \frac{1}{2} \log_e (3.6269 + 3.7622) \\ &= \frac{1}{2} \log_e (7.3891) = \frac{1}{2} \text{ of } 2 = 1 \end{aligned}$$

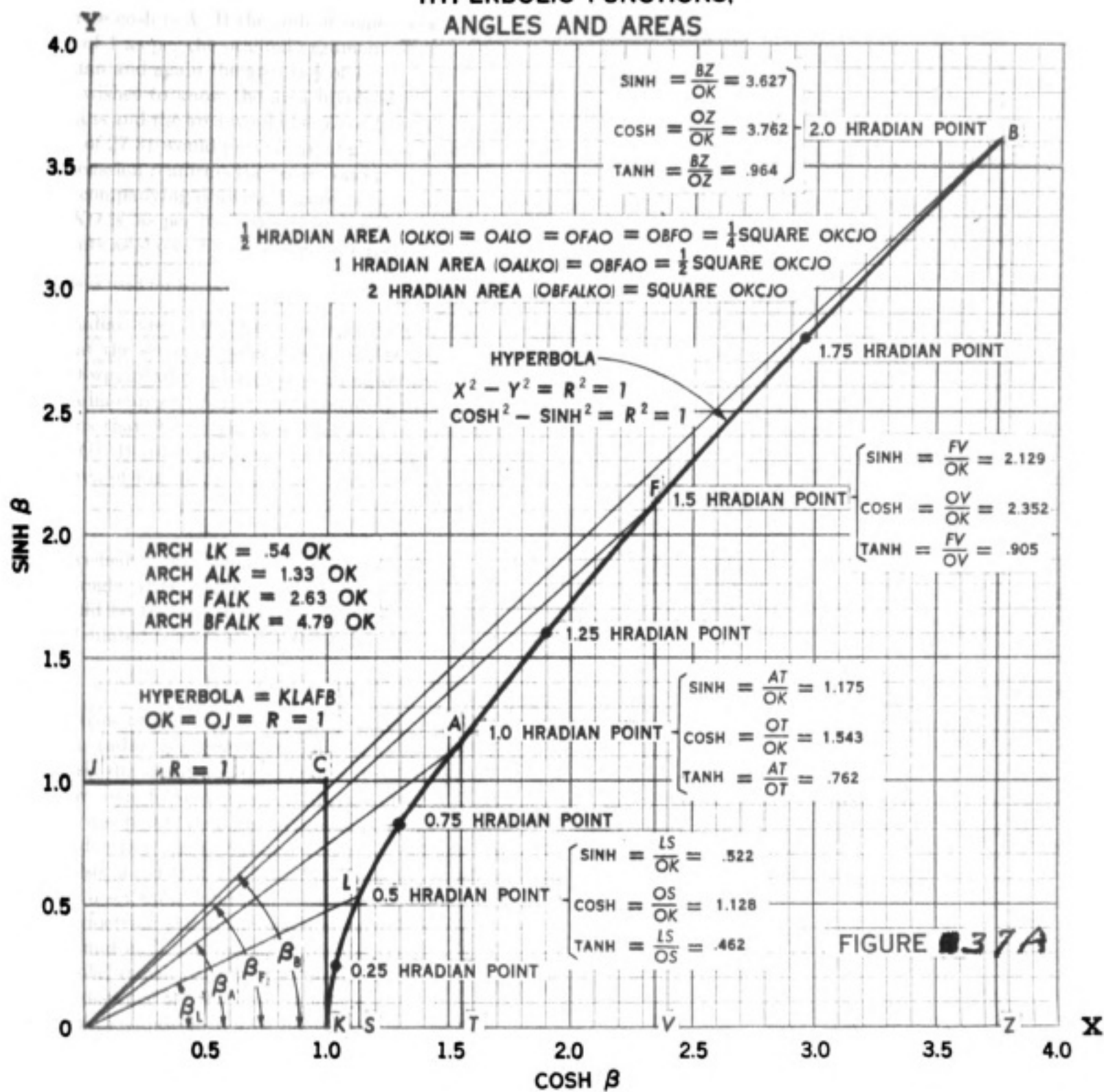
In the above equations the letter e is used as a matter of convenience instead of the Greek letter ϵ , sometimes used

The log of 7.3891 to the base e as per table is 2. In other words, $e^2 = 7.3891$. The area is one square inch which is two unit areas.

There is a check in these data. The doctrine of hradians, based on half square inch units, is again exemplified and validated.



HYPERBOLIC FUNCTIONS, ANGLES AND AREAS



The above formula for the area in question, namely $\frac{1}{2} \log_e (X + Y)$ can be written readily in another form, namely

$$\text{Area} = \frac{1}{2} \sinh^{-1} Y = \frac{1}{2} \cosh^{-1} X.$$

This equation says that area equals $\frac{1}{2}$ the hyperbolic angle whose $\sinh$ is Y and also equals $\frac{1}{2}$ the hyperbolic angle whose $\cosh$ is X . If the $\sinh$ of some hangle is 1.1752, then the angle is *one* hyperbolic radian and so the area is $\frac{1}{2}$ of 1 as per the original argument. If the $\cosh$ of some hangle is 1.5431, then the hangle is *one* hyperbolic radian and again the area is $\frac{1}{2}$ of 1.

If one wishes to know the area involved for a hyperbolic angle whose $\sinh$ is 2.1293, the hangle would be 1.5 hradians and the area would be .75 or $\frac{1}{2}$ of 1.5. (1.5 unit areas)

A $\cosh$ of 27.31 would provide an angle of 4 hradians and an area of 2 (4 unit areas) would be involved. This latter discussion reaffirms the basic concept on which hyperbolic radians was defined.

The accompanying drawing Figure 37 shows areas corresponding to 1, 2 and 2.5 hradians. The first hradian area $OAKO$ is, as per basic doctrine, equal to one-half the square area, $1DKO1$. If the square area were 100 square units as shown by cross-section paper, then hradian area would be 50 square units.

hradians is used for hyperbolic radians and hangle is used for hyperbolic angle.

The hradian area, $OBKO$, corresponding to two hradians, is equal in area to the square $1DKO1$. This means that the second hradian area, $OBAO$, is, as marked, one-half of the square area. It is one-half square unit or 50 square units in area depending on the scale of areas used in drawing.

The hradian area $OCBAKO$ corresponding to two and one-half hradians is 1.25 times the square area, $1DKO1$. This means that the long and slender area in solid black, $OCBO$, in the drawing, is one-fourth of the square area $1DKO1$. It corresponds to one-half hradian. It is interesting to compare this needle shaped area to the .5 hradian area of Fig. 37.

AREA UNDER CURVE

If one wished to find the area (AGK) under the curve, it would be easy to determine. The area of the right angled triangle AGO would be .5 XY . Therefore, .5 XY (.5 $AG \cdot OG$) minus the appropriate area as discussed above would be the area under the curve for any point X, Y on the curve. Areas BEK and CFK would be readily calculated from triangles BEO and CFO .

HYPERBOLIC ARC AS A POSSIBLE DEFINITIVE SPECIFICATION

The doctrine of defining one circular radian of angle as that angle which is subtended by an arc of the circle equal to the radius is a simple, straight-forward and tenable doctrine. The area of the sector as .5 square inches as previously discussed was the important result of this procedure.

In the previous discussion page 76, Figure 36 of how one hyperbolic radian was defined the arc doctrine cannot be used. It was pointed out the .5 square inch area doctrine was used as a primary and fundamental approach. This same doctrine was the primary basis for defining a circular radian.

It was also noted that the "arc" doctrine (a secondary approach) which worked for the circle might have been tried to set up a valid and tenable definition of an hyperbolic radian.

The arc doctrine is satisfactory and tenable for the circle. The reader after understanding the discussion about $\sinh$ and $\cosh$ will readily subscribe to the idea that the arc doctrine could not be used in the case of the hyperbola. It will be readily appreciated that the arc or distance along the hyperbola curve would not be a constant distance for the individual successive hyperbolic radians.

The equal .5 square inch area doctrine is the only valid and tenable doctrine for defining both the circular radian and the hyperbolic radian.

LENGTH OF HYPERBOLIC ARC

It would be interesting to ask how long is the arc AK in sketch along the hyperbola shown for the first hradian and BAK for the first two hradians, Figure 37.

Here again, one asks the expert mathematician for an "assist". Even though the equation of the equilateral or rectangular hyperbola is a simple one, namely $X^2 - Y^2 = 1$ and even though the general technique of finding length of arc by integral calculus is a fairly simple and straightforward technique, the determination of the length of arc in this particular case happens to be a bit complicated and calls for great skill in mathematics as such.

However, the expert mathematician comes up with a fairly simple formula. The numerical value for length of arc is given approximately by the following simple equation. Length of Arc $= \frac{2}{3}Y - \frac{1}{2} \tan^{-1} Y$.

Here Y is the ordinate (sinh) of the particular point in question. The second part of the equation ($-\frac{1}{2} \tan^{-1} Y$) says, minus one half the angle in circular radians whose tangent is Y .

SOME NUMERICAL VALUES OF HYPERBOLIC ARC

The following values of arc were calculated by the formula above.

For 1 hradian, whose Y (numerically the sinh 1) is 1.1752, the arc AK is 1.330

$$1.330 = \frac{2}{3}(1.175) - \frac{1}{2} \cdot .8656$$

For 2 radians, whose Y (numerically the sinh 2) is 3.6269, the arc BK is 4.790

$$4.790 = \frac{2}{3}(3.6269) - \frac{1}{2} \cdot 1.3$$

For 2.5 radians, the arc $CK = 8.3728$

$$8.3728 = \frac{2}{3}(6.0502) - \frac{1}{2} \cdot 1.405$$

If one calculated the straight line distance AK for one hradian it would come to be 1.300. This distance would be the hypotenuse of a right triangle whose sides are 1.175 and $(1.543 - 1)$. The arc AP is a bit longer; namely 1.330 as compared to 1.300.

For two hradians, the straight line distance BK , is 4.550. The sides of the right triangle in this case would be 3.63 and $(3.76 - 1)$. The arc is a bit longer, namely 4.790 as compared to 4.550.

In both cases, the straight line distance is the square root of $[\text{Sinh}^2 + (\text{Cosh} - 1)^2]$

As previously stated, the formula given for length of arc provides an excellent approximation for length of arc. A completely rigorous analysis provides a more precisely correct value for the 1 hradian arc as 1.317 in contrast to the above value of 1.330. For two hradians, the more exact value for the arc is 4.625 in contrast to 4.790. The approximate values are a bit on the "high" side as compared to the more rigorously calculated values. The approximate formula is good to have however since it provides an easy determination of the arc length.

Finally it is to be seen, that the several arcs for the consecutive individual hradians are not constant in length as they were for circular radians but have a continuously increasing length as given on previous page.

It can be observed by inspection of accompanying sketch that the individual arcs for the .1 circular radians are constant in length but the individual arcs for the .1 hyperbolic radians increase a tiny bit in length.

The areas involved for the first, second, third, etc circular radians and hyperbolic radians are all the same, namely .5 square inch. For the tenth's of angle, the area in each case is .05 square inch.

ASYMPTOTE AT 45° ANGLE

The asymptote is shown in the sketch and since the hyperbola is an equilateral or rectangular one and one asymptote goes out straight northeast at 45° to the X axis. At infinity, the hyperbola touches the asymptote, so even though there would be millions of hradians created, their total sum from the circular angle point of view would add up to only 45° circular degrees. From the point of view of circular angles, an infinite number of hradians is only 10 circular degrees greater than the circular degrees of 1 hradian ($45^\circ - 35^\circ$). The circular angularity of the later individual hradians is very small and get smaller and smaller but they all provide .5 sq. inches areas. They get very long and "skinny" in shape but the area provided in each case is the same, namely .5 square inches.

GRAPHIC REPRESENTATION OF CIRCULAR AND HYPERBOLIC SINES, COSINES AND TANGENTS

In the accompanying drawing previously referred to in connection with circular radians and hradians, the radius (R) of the circle is purposely made equal to " a " of the hyperbola. Furthermore, the " R " and the " a " are assumed to be numerically equal to unity (1). The circular angle B is purposely made one circular radian and the hyperbolic angle is one radian. All circular radians are each 57.296° , but number one hyperbolic radian is 35 circular degrees but number 2 hradian has fewer circular degrees, etc. The expression of hyperbolic angles in circular degrees is of no importance but does aid a bit in understanding and appreciation of the problems involved in hyperbolic functions as contrasted to circular functions.

Inspection of the drawing will readily graphically reveal the relative numerical values of (indicated graphically by length of line) $\sin \alpha$ and $\sinh \beta$, $\cos \alpha$ and $\cosh \beta$.

$$\sin \alpha = \frac{AC}{AO} = \frac{AC}{1} = AC \text{ (graphically)}$$

$$\cos \alpha = \frac{OC}{AO} = \frac{OC}{1} = OC \text{ (graphically)}$$

Take a look at sketch, Figure 36.

$$\sinh \beta = \frac{AD}{MO} = \frac{AD}{1} = AD \text{ (graphically)}$$

$$\cosh \beta = \frac{OD}{OK} = \frac{OD}{1} = OD \text{ (graphically)}$$

Take a look at sketch, Figure 36

In the drawing of the equilateral or rectangular hyperbola, page 68 Figure 32 it will be observed that the asymptotes go off in space to the northeast and southeast at 45 degrees, which is a characteristic of this particular curve. The northeast asymptote has the equation, $X = Y$. The southeast asymptote has the equation, $X = -Y$.

The eccentricity is in this case of the equilateral or rectangular hyperbola, as pointed out before, equal to

$$\sqrt{2} = \frac{\sqrt{a^2 + a^2}}{a} = \frac{\sqrt{b^2 + b^2}}{b} = 1.41$$

provided

$$a = b. = 1$$

PRECISE NUMERICAL VALUES OF CIRCULAR SINES AND COSINES

Obviously one could find the approximate numerical values of circular sines and cosines of given angles by actually drawing a circle with a compass and measuring the various angles with a protractor (a sort of angular ruler), then actually making right triangles with a protractor and ruler and then finally measuring with a divider and ruler the magnitude of the length of hypotenuse, side opposite and side adjacent (the X or abscissa).

If the above could be done accurately, and of course it could not, then laborious arithmetic would have to be involved finally to get the ratios for the sines, cosines and tangents for a whole lot of angles. Instead of all this, we can ask again for help from the expert mathematician.

MATHEMATICAL EXPANSION SERIES

We have the basic equation of the circle

$$X^2 + Y^2 = 1 = \cos^2 \theta + \sin^2 \theta$$

What must be done to find $\frac{Y}{R} (\sin \theta)$ and $\frac{X}{R} (\cos \theta)$ with precision as to numerical values, is to solve the equation in an overall general manner for all combinations of X and Y .

This means, what values of $\left(\frac{X}{R}\right)^2$ and $\left(\frac{Y}{R}\right)^2$ for every point and for every angle will make $\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1$, or in other words, will satisfy the equation.

We ask again for an assist from the mathematics expert. He states that in the first place, the angle cannot be talked about in degrees (a perfectly conventional and arbitrary method) but must be expressed in radians.

There are 2π radians of angle in a whole circle. It happens that one radian is $\frac{360}{2\pi}$ or 57 degrees, 17 minutes 45 seconds, but this value is a part of the conventional assignment of 360° in a circle but neither the 360 nor the 57 plays a basic and fundamental role in the whole endeavor. It (360) is a good practical number for general use because it can be divided by so many smaller numbers to give a whole number answer.

So with his beautiful mathematics, which is really neither too complicated nor beyond the understanding of anyone who will give it a good try, involving converging expansion series and certain basic theorems by McLaurin and Taylor, he provides us the following relatively simple series or array of terms which will give the precise numerical values of $\sin \theta$ and $\cos \theta$ which will satisfy the basic equation.

$$\text{The sine of } \theta \text{ (radians)} = \theta - \frac{\theta^3}{|3|} + \frac{\theta^5}{|5|} - \frac{\theta^7}{|7|} + \cdots$$

$$|3| \text{ means factorial } 3 = 1 \cdot 2 \cdot 3 = 6$$

$$|5| \text{ means factorial } 5 = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

$$|7| \text{ means factorial } 7 = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$$

This expression of a function, as for example $\sin \theta$, as the sum of a series of terms which is easily available numerically, is a great and valuable "tool".

So to illustrate how nicely this series of terms gives the numerical value of a sine of an angle very accurately and relatively very easily, assume that the angle is 1 radian.

$$\begin{aligned} \text{So sin 1 radian} &= 1 - \frac{1}{6} + \frac{1}{120} - \frac{1}{5040} \\ &= 1 - .16666 + .00833 - .00020 \\ &= 1.00833 - .16686 \\ &= .84147 \end{aligned}$$

If the sine of one radian ($57^\circ 18'$) is looked up in a table of natural sines, it will be given as .84147.

If the reader wishes to go through with the final steps in arithmetic for .5 circular radian ($28^\circ 39'$), he will have the fun of finding that the numerical value of the series expression, for $\sin .5 \text{ (radian)} = .5 - \frac{.5^3}{6} + \frac{.5^5}{120} - \frac{.5^7}{5040}$ will come out as .47943, which is the value given in a table.

It would be of interest to find the sine of 2 radians which is an angle of $114^\circ 36'$. The sine of $114^\circ 36'$ is of course equal to the cosine of $24^\circ 36'$.

By using 2 radians for θ in the series expansion expression, one obtains

$$\begin{aligned} \sin 2 \text{ radians} &= 2 - \frac{8}{6} + \frac{32}{120} - \frac{128}{5040} \\ &= 2.266666 - 1.358729 \\ &= .907937 \end{aligned}$$

This value, the result of only four terms in the series expression, is close to the cosine of $24^{\circ}36'$ as given in the table.

One knows from a consideration of the basic nature of the sine and cosine that their maximum value is unity (1). Whatever large number of radians are used in the series expression, the numerical result never exceeds unity (1). The second term $-\frac{\theta^2}{2}$ takes care of the situation for the sine. Its minus value will always be greater than 1 for values of θ greater than $\sqrt{6}$, namely 1.817.

Of course the second term gets an assist from the fourth term.

Of course to find the sines of many angles in radians and fractions or multiples thereof corresponding to so many degrees by use of the series expansion expression took a terrific amount of work on somebody's part. It is however the only relatively easy and valid method to find the precise numerical value to several decimal places.

SERIES EXPANSION FOR COSINE

By the same fine mathematical process, the numerical value of $\cos \theta$, where θ the angle must be known in radians, can be found by the expansion series,

$$\cos \theta = 1 - \frac{\theta^2}{2} + \frac{\theta^4}{24} - \frac{\theta^6}{720} \pm \dots$$

Maybe some reader would like to check from the table of natural circular cosines in some book, the numerical value of cosine θ for 1 radian by his own arithmetic; also one might check for 2 radians.

The tangent expression series is by way of completion of matters

$$\tan \theta \text{ (radians)} = \theta + \frac{\theta^3}{3} + \frac{2\theta^5}{15} + \frac{17\theta^7}{315}$$

The values in those denominators are not factorial values.

Actually of course since $\tan \theta = \frac{\sin \theta}{\cos \theta}$, the numerical value of a tangent can be more readily and precisely found by using the simple ratio $\sin \theta$ and cosine θ values and not do a lot of arithmetic as required in above series expansion.

The circular sine and circular cosine curves shown in an earlier sketch (page 27) show the periodic or recurring values of sines and cosines as the ordinate or Y distances of the curves. These increase and decrease in numerical value. The abscissa of these curves are the various increasing values of the angle itself in either circular radians or degrees as the radius swoops around in counter clockwise rotation at time increases.

SINH, COSH, EXPANSION SERIES, PRECISE NUMERICAL VALUES

The same argument and technique which was used to find the precise numerical evaluation of circular sines and circular cosines, can also be used to find the numerical value of sinh and cosh. The question is again, what values of $\left(\frac{X}{a}\right)^2$ and $\left(\frac{Y}{a}\right)^2$ for every point on the hyperbolic curve will satisfy the equation $\frac{X^2}{a^2} - \frac{Y^2}{a^2} = 1$, where "a" is assumed to be unity.

Again it is to be remembered that sinh and cosh are not periodic or recurring factors. Sinh will start off at a certain value for number 1 angle measured in hradians and then continuously increase as the route "goes way off in the wide blue yonder". The numerical value of cosh will do the same as will the tanh. The value of sinh and cosh become very large but tanh grows only to unity as the hyperbolic angle increases. See Fig. 38

Whatever is true relative to the numerical values of sinh and cosh, their numerical value is imperiously decreed by the equation $X^2 - Y^2 = 1$ for an equilateral or rectangular hyperbola which is the particular hyperbola on which hyperbolic functions are based.

The splendid mathematical principle of a function being expressible as a summation of a series of terms

again makes an "assist", as they say in baseball to provide with great precision the numerical values of sinh, cosh and tanh.

$$\text{Sinh } \beta \text{ (hradians)} = \beta + \frac{\beta^3}{3} + \frac{\beta^5}{5} + \frac{\beta^7}{7} + \dots$$

$$\text{Cosh } \beta \text{ (hradians)} = 1 + \frac{\beta^2}{2} + \frac{\beta^4}{4} + \frac{\beta^6}{6} + \dots$$

$$\text{Tanh } \beta \text{ (hradians)} = \beta - \frac{\beta^3}{3} + \frac{2\beta^5}{15} - \frac{17\beta^7}{315} \pm \dots$$

On page 82, the calculation for $\sin \theta$ where θ was 1 circular radian was illustrated. It will be observed that the series expression for hyperbolic sine (sinh) is the same as for the circular sine except as to signs. The signs of the terms in the sinh and cosh series are all plus, meaning that as β increases, the cosh β and the sinh β continuously increase.

$$\begin{aligned} \text{So sinh 1 radian} &= 1 + .16666 + .00833 + .00020 \\ &= 1.17519; \text{ the value given in the table.} \end{aligned}$$

Maybe some reader would like to calculate the numerical value of cosh β for 1 hradian from the series expansion, by his own arithmetic, and then check with the value given in a table.

RANGE OF NUMERICAL VALUES

The maximum possible numerical value of a circular sine and cosine is unity (1). The maximum numerical value of a circular tangent is infinity.

The maximum numerical value of sinh and cosh is infinity. The maximum value of tanh β is unity (1).

Furthermore in contrasting the two kinds of functions, it is to be noted that the numerical value of the circular sine is zero under the circumstances when the numerical value of circular cosine is unity (1). When the circular cosine is zero, the circular sine is unity (1). It is to be noted that the hyperbolic sine is also zero when the cosh is unity (1). However, the numerical value of cosh is never zero. Its minimum numerical value is unity (1). As the hyperbolic angle β increases from zero, not to a maximum of 2π radians as in the case of the circle, but to a maximum of infinity radians, the sinh and cosh become almost equal in value relatively soon and keep on increasing in equal value and attain infinity as a maximum value.

Simple and graphical analysis and inspection of the geometric and trigonometric foundations of circular sine, cosine and tangent and of hyperbolic sine, cosine and tangent readily disclose the interesting contrasts in numerical values above mentioned.

GRAPH OF VALUES OF SINH, COSH AND TANH

The accompanying chart Figure 38 illustrates what has been pointed out in previous pages about how the numerical values of sinh, cosh and tanh vary with increasing number of radians.

At the beginning, up to only 1 hradian, when the hyperbola itself is curving a lot, all three functions change fairly rapidly; their curves bend. After a while the sinh and cosh curves continue to change (but bend only a little) and have an almost constant slope. The numerical value of sinh and cosh continuously increase to a maximum value of infinity. The tanh curve levels off and becomes flat. After 4 hradians, sinh and cosh become almost equal in numerical value. This is reasonable since the hyperbola itself almost becomes a straight line with 45° of slope and nearly parallels the asymptote. The Y (sinh) and the X (cosh) become nearly equal since X and Y become nearly equal far out on the hyperbola. After a while, the equation of the hyperbola almost become $Y = X$ which is the equation of the asymptote straight line for 45° .

The tanh curve $\left(\frac{Y}{X}\right)$ becomes practically flat because out on the hyperbola, Y becomes practically equal to X . The tanh $\left(\frac{Y}{X}\right)$ nears unity at 4.2 radians.

HYPERBOLIC SINE, COSINE AND TANGENT CURVES

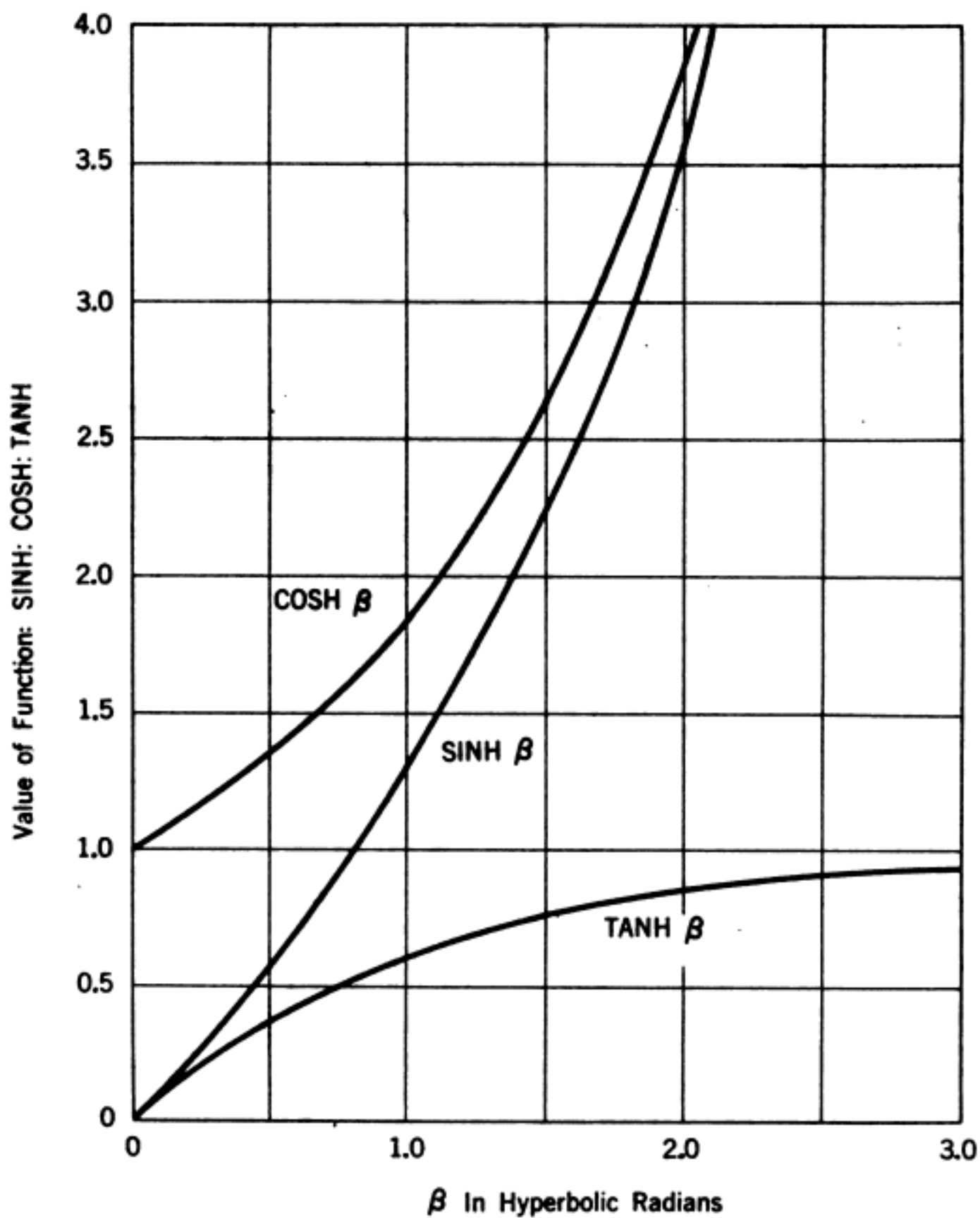


FIGURE 38

SINE, COSINE AND TANGENT CURVES

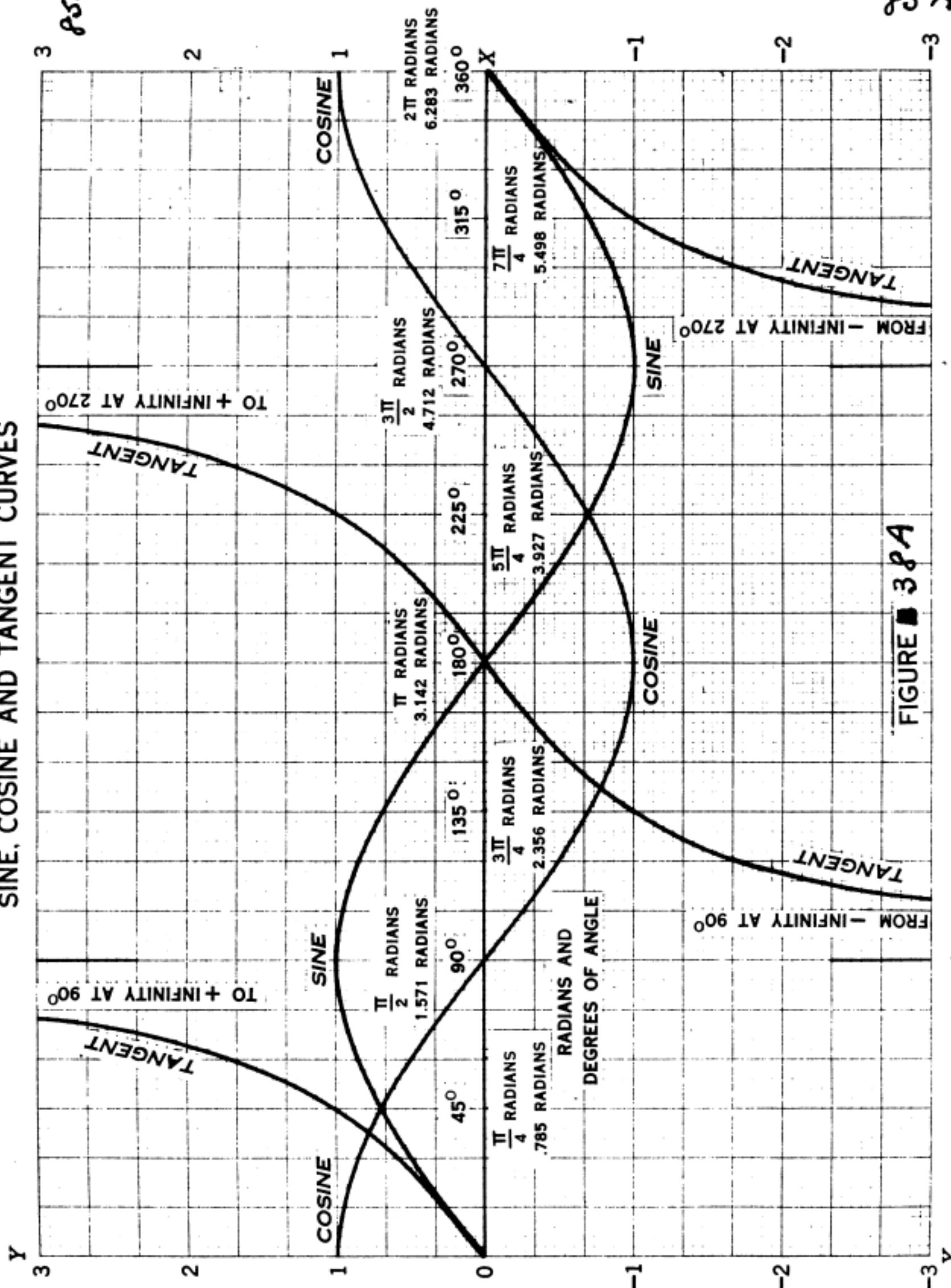


FIGURE 38A

85A

85A

It is suggested that the reader take a good look at a table of sinh and cosh in some book of tables. The tables may not look exciting but examination of them will reveal some interesting facts and will promote understanding.

PRECISE NUMERICAL VALUE OF π

In connection with the series expansion doctrine, the reader might be interested in the series expansion for π , the ratio of a circumference to the diameter of a circle. Its numerical value could not be possibly determined very precisely by sheer measurement with a ruler, divider and piece of string

$$\begin{aligned}\frac{\pi}{4} &= \frac{1}{2} - \frac{1}{3 \cdot 2^3} + \frac{1}{5 \cdot 2^5} - \frac{1}{7 \cdot 2^7} \pm \dots \\ &+ \frac{1}{3} - \frac{1}{3 \cdot 3^3} + \frac{1}{5 \cdot 3^5} - \frac{1}{7 \cdot 3^7} \pm \dots\end{aligned}$$

By using only 9 terms of the first group of terms in the above and only 5 terms of the second group

$$\frac{\pi}{4} = 0.463647 + 0.321751$$

$$\pi = 3.141592$$

The numerical value of π is incommensurate. It has no equivalent as a pure fraction as for example 22/7. Its digits do not repeat themselves. To check this fact, the value of π has actually been carried to 200 decimal places by specialists. To fifteen decimal places, $\pi = 3.14159\ 26535\ 89793$. One observes that there is no repetition of digits.

The value of 22/7 (a definite ratio) is 3.142857 142857 142857 and has a repetition of digits.

CONCLUSION

Oftentimes as said previously students regard hyperbolic functions (sinh, cosh and tanh) as very complex, abstruse and "high brow". This indictment is not valid. They can be readily and nicely understood and appreciated, if they are approached in a simple and step-by-step fashion with graphical illustration and interpretation. This approach has been used in the present treatise. Of course a departure from one's circular background and concepts must be made.

It is possible that the reader may think that the expository and narrative style herein used is repetitious and a bit redundant. The defense to this indictment is the writer's conviction that the learning process is not a "one shot" affair. Frequent reviews of ideas already grasped and understood, different approaches, and a gradual and logical arrival at the threshold of new ideas and concepts seem to be the requirement of ordinary mortals.

The great generic ideas and principles, now accepted as basic truths in the physical sciences, had a very long period of gestation. Furthermore it takes a long time for ideas to grow up. Ideas and techniques have to be matured so very long in the cradle of time before they "step out" into practical endeavors.

The path to understanding and to skill and competence is not easy, is not a super highway and is not a royal road. It can be traveled however if one gives it a good try.

CHAPTER 7

Some Fundamental Considerations of Conic Section Curves as Created by Intersecting Plane and Cone

It is suggested to the reader that he take several looks at the first photographs page 8 before he reads and studies the following discussion. Some inspection of and reflection about the geometric portrayals in the photographs will provide a good basis for study.

Consider, as indicated in accompanying drawing, Figure 39, a plane passing through the apex (A) of a right circular cone and intersecting base of cone (BC) at right angles (90°). An isosceles triangle would result from the intersection. The top angle of the triangle will be called θ (theta) and the equal angles at the base will be called β (beta). The angle β is really the base angle of the cone itself.

CIRCLES

If a plane intersects a cone at right angles to the vertical axis (AD) of the cone, or, in other words, if the intersecting plane is parallel to the base of cone BC , a circle would be created. If the intersecting horizontal plane is varied in vertical distance from base line BC , a family of circles would result. The circles would be of the same shape but of different size.

The eccentricity of all the circles would be zero (0).

There is only one possible angle of intersection to provide a circle.

PARABOLAS

If the intersecting plane cuts the cone at a certain angle α (alpha) to the base line BC , or, in other words, if the intersecting plane is parallel to the slant surface of cone AB or AC , a parabola is created. This means that $\alpha = \beta$. If the intersecting plane is varied in distance along slant side AC of cone, a family of parabolas would result. The parabolas would be the same shape but of different size.

The eccentricity of all the parabolas would be unity (1).

There is only one possible angle of intersection to provide a parabola.

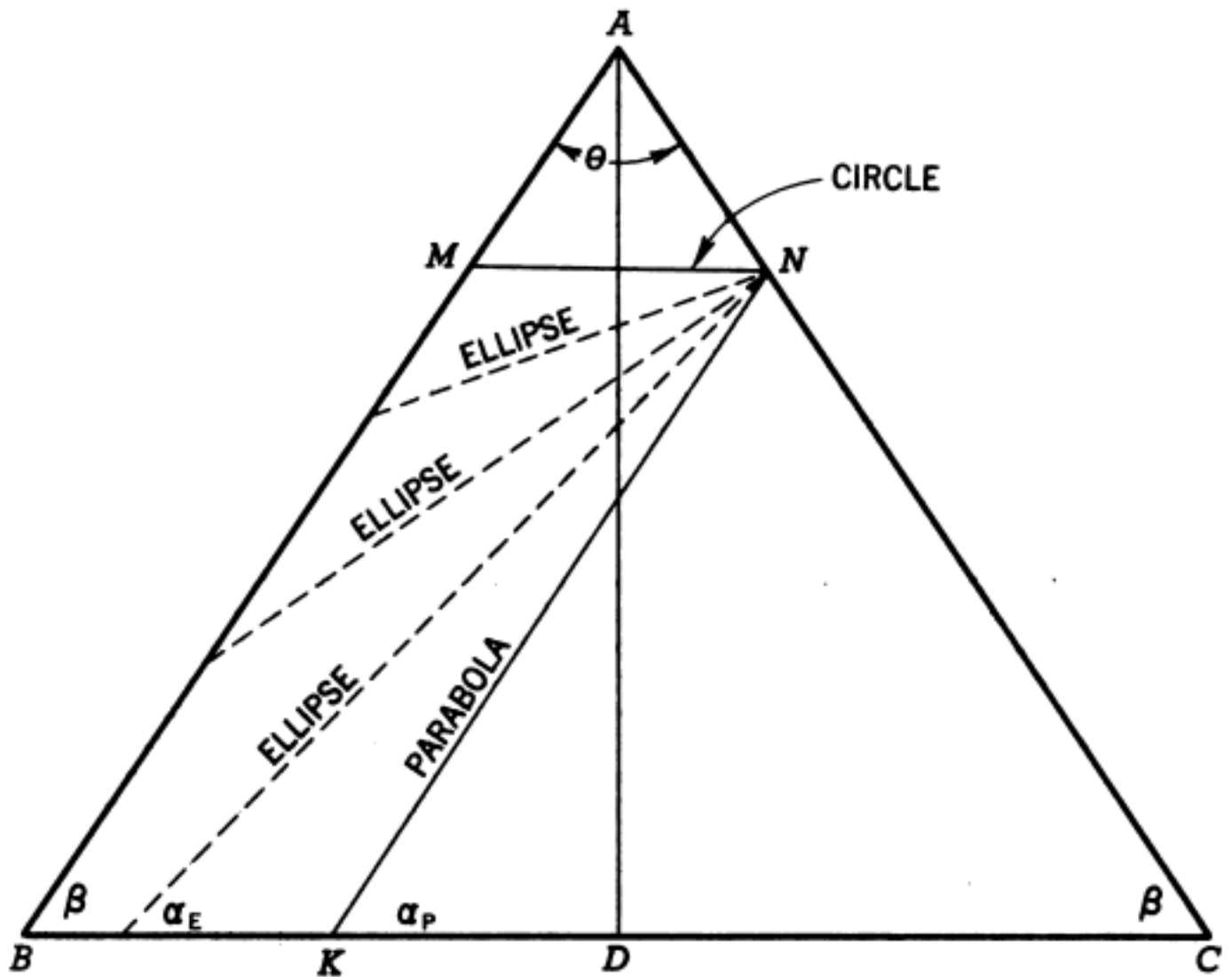
In drawing, the angle of line in intersecting plane with base line of cone is called α (alpha.) When for any position of the intersecting plane $\alpha = \beta$, a family of parabolas result.

ELLIPSES

If a line in the intersecting plane cuts the base line of cone at some angle less than β , then a family of ellipses would result. This means that the angle α is less than angle β . The eccentricity of all the ellipses is less than unity (1). The ellipses are of different shape and size. There are many possible angles of intersection to provide an ellipse.

For some additional analysis, consider accompanying drawing, Figure 40 further. The line MN is the line of the circle. The angle MNK is α_p , the parabola angle and also the angle, β . The line NK is to be rotated clockwise about point N . In position NK , a parabola results. But as line NK is rotated, the angle α_s decreases to angle zero (0) when NK becomes coincident with line NM . For all the angles from α_p to zero (0), ellipses are created. When angle α_s (MNK) becomes 0, then a circle results. At point N , there are three ellipse angles, α_s , α_{s1} and α_{s2} .

To get the complete ellipse for angle α_s , the dotted line in intersecting plane must emerge from the slant side line of cone, AB . The cone must be imagined and the isosceles triangle must be imagined extended farther downward to the "southwest" than shown in Fig. 2, page 9.



For all Circles, $\alpha = 0 \quad e = 0 = \frac{\sin 0}{\sin \beta}$

For all Parabolas, $\alpha = \beta \quad e = 1 = \frac{\sin \alpha}{\sin \beta}$

For all Ellipses, α varies from β to 0

Eccentricities vary as $\frac{\sin \alpha}{\sin \beta}$ vary = Less than 1

FIGURE 40

THE ELLIPSE AND A CYLINDER

If a plane should cut a cylinder parallel to its base, the curve produced would be a circle. The angle β of the cylinder would be 90° and the angle α would be 0° . If the angle α would be of some value greater than 0° , then the curve produced by the intersection of the plane and the cylinder would be an ellipse.

The validity of the latter statement follows from the idea that a cylinder can be regarded as a cone of infinite height: a cone whose angle β is 90° . Applying the general criterion for eccentricity, $e = \frac{\sin \alpha}{\sin \beta}$, the e in this case would be $\frac{\sin \alpha}{\sin 90} = \sin \alpha$ since $\sin 90^\circ = 1$. The e is less than unity (1) and the conic curve is therefore an ellipse.

$$\text{ECCENTRICITY OF CONIC CURVES} = \frac{\sin \alpha}{\sin \beta}$$

To guide and assist one's understanding of the conic section curves, it is in order to consider the fortunate and important fact that the eccentricity of the conic curves is simply related to the trigonometric sines of the angles α and β . The simple relation is

$$\frac{\sin \alpha}{\sin \beta} = \text{eccentricity} = e$$

The niceties and the particulars of the proof of this relation are not difficult. The reader is referred to a text-book on analytical geometry for the proof. In the text "An Elementary Course in Analytical Geometry" by J. H. Tanner and Joseph Allen, published by The American Book Company in 1898, a splendid and straightforward development of this simple and important relation is given in the Appendix, Note D, pages 384-386. A helpful three-dimensional drawing is used. Of course this proof is found in more recent books too.

For the purposes of this treatise, the fortunate and simple relation will be accepted and used as a fundamental fact and as a "tool." The work and skill of the expert geometrician provides a fine "assist" again.

A bit of review in the light of the above relation is in order. For all the circles, $\alpha = 0$. Therefore,

$$\frac{\sin 0}{\sin \beta} = \frac{0}{\sin \beta} = 0 = \text{eccentricity} = e.$$

For all the parabolas, $\alpha = \beta$. Therefore,

$$\frac{\sin \alpha}{\sin \beta} = 1 = \text{eccentricity} = e.$$

For all the ellipses, α is less than β . Therefore,

$$\frac{\sin \alpha}{\sin \beta} \text{ is less than unity (1) } = \text{eccentricity} = e.$$

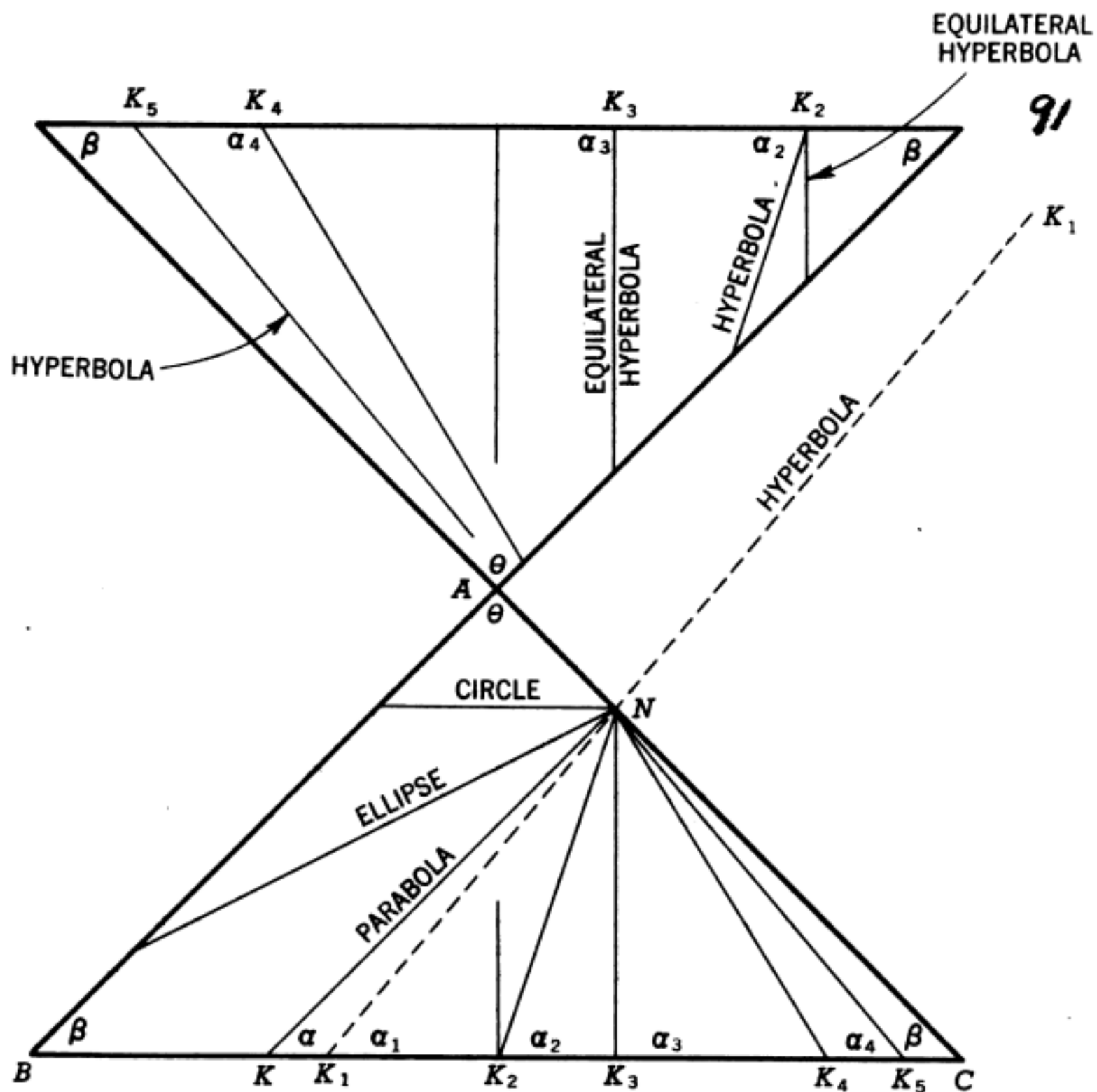
HYPERBOLAS

One of the principal objectives of this treatise is an understanding and rationalization of the hyperbolic functions $\sinh$ and $\cosh$ as such on the one hand and their use in electrical communication matters on the other hand. This latter rationalization is done in later pages. It is in order now to give some particular attention to the hyperbola as such to the hyperbola as a conic section curve.

The sketch of accompanying drawing, Figure 41 portrays a vertical plane section through common vertex (point A) of two cones.

The line NK whose position provides the parabola is now to be rotated counter clockwise. The first small rotation is indicated by dotted line NK_1 . If dotted line NK_1 were extended to the northeast as it were, it would intersect the upper cone if upper cone were extended. The upper and lower cones are oftentimes called the upper and lower nappes.

The discussion to follow is concerned with the important and simple idea that the lines NK_1 , NK_2 , NK_3



HYPERBOLAS

$$\theta = 90^\circ \quad \beta = 45^\circ$$

For all Hyperbolas $\alpha > \beta \quad e > 1 = \frac{\sin \alpha}{\sin \beta}$

Eccentricities vary as $\frac{\sin \alpha}{\sin \beta}$ vary $= > 1$

Equilateral Hyperbolas $= \frac{\sin 90^\circ}{\sin 45^\circ} \quad e = 1.41$

FIGURE 41

NK_4 and NK_5 in the intersecting plane create hyperbolas. These lines are really the near edges of the conic curves created by the intersecting plane and the surface of the cone. To see, in one's imagination, the hyperbolic curves, one would need to look from a point way out to the east or west.

Since the angles of intersection $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and α_5 are greater than α and also greater than β , hyperbolas are created in both upper and lower nappes.

For the cases of parabola, ellipse and circle when $\alpha = \beta$ and when α varies from α to zero (0), the intersecting plane cannot cut both nappes.

It is to be noted again that the angles α_1, α_2 etc. are greater than α and β . In fact the criterion for the creation of hyperbola is that the angle β (the angle of intersection formed by base line and rotating line with N as center) must be greater than the angle specified for creation of a parabola (β), the base angle of the cone.

Consider the extension of line NK_2 extended to northeast to base line of upper cone. An equal angle α_2 is indicated at the upper base line.

Now as per basic doctrine previously mentioned,

$$\frac{\sin \alpha}{\sin \beta} = \text{eccentricity} = e$$

Since the α angles are greater than β , the eccentricity for all the hyperbolas is greater than unity as per basic criterion of hyperbolas.

The eccentricities of the hyperbolas in upper and lower nappes, as the result of the same intersecting line, are of course equal since the angles of intersection at base of upper nappes are equal to the angles at base of lower nappe. As a matter of interest, it is suggested that the sketch be turned upside down.

The hyperbolas of the upper and lower nappes are not outlined or created to the same amount. For angle α_1 , formed by dotted line, the lower nappe hyperbola would be outlined quite a bit along its path while the upper hyperbola way up to the northeast beyond the page would be outlined to a much less extent. For angle α_2 , the lower hyperbola would be more extensively created than the upper "twin". For angle α_4 , the upper "twin" would be created to a greater degree along its path than the lower "twin".

It is to be restated that the eccentricity for upper and lower "twin" hyperbolas of all the hyperbolas are equal and greater than unity (1).

Angle $\alpha = 90^\circ$, A Special Case

As is to be observed, the angle α_2 is 90° . Therefore when the intersecting line in the plane is parallel with common vertical axis, the upper hyperbolas have the maximum eccentricity.

This is true because the numerator, $\sin \alpha$, has its greatest value of unity (1)

$$\frac{\sin \alpha_2}{\sin \beta} = \frac{\sin 90}{\sin \beta} = \frac{1}{\sin \beta}$$

The two hyperbolas in this case are created to the same degree along their path.

If β were assumed to be 30° , the eccentricities of the "twin" hyperbolas in the case at hand would vary from near $\frac{.5}{.5}$ or 1 to $\frac{1}{.5}$ or 2 and then back near to 1, as line NK was rotated counterclockwise to parallelism with slant side, AC .

SLENDER CONES AND HYPERBOLAS

If one had a very slender cone with an apex angle θ , a small angle (4°) and a base angle a large angle of 88° , then all the hyperbolas would have eccentricities only a bit greater than 1. All of these hyperbolas would be very near parabolas.

The hyperbolas created in this case when the intersecting line in plane is parallel to vertical axis would have maximum e 's of 1.0006

$$\frac{\sin \alpha}{\sin \beta} = \frac{\sin 90^\circ}{\sin 88^\circ} = \frac{1}{.9994} = 1.0006$$

The eccentricity of the parabolas would be

$$\frac{\sin \alpha}{\sin \beta} = \frac{\sin 88^\circ}{\sin 88^\circ} = \frac{.9994}{.9994} = 1$$

THE EQUILATERAL HYPERBOLA IS OF SPECIAL IMPORTANCE

If the cone had a 90° angle (θ) at the top, then the angles β would be 45° . Under these circumstances, the hyperbolas created, when the intersecting plane is parallel to vertical axis, would have eccentricities of 1.41

$$\frac{\sin \alpha}{\sin \beta} = \frac{\sin 90^\circ}{\sin 45^\circ} = \frac{1}{.707} = 1.41 = e$$

The interesting and important fact is thus brought out, that the equilateral or rectangular hyperbola whose eccentricity is 1.41 is created in the above special case.

To see this equilateral hyperbola, one must imagine he is looking west from a point far to the east or vice versa. See photograph, Figure 9.

As has been stated elsewhere in this treatise, the equation of the equilateral hyperbola is

$$X^2 - Y^2 = 1$$

This particular hyperbola is of special importance and significance because it is the basis of hyperbolic sines and cosines. These were discussed, developed and rationalized in previous pages.

For the case of the equilateral or rectangular hyperbola, it is easy to imagine that the slant lines of the conic surfaces are the asymptotes of the curve. These asymptotes go out at 45° to the "northeast" and "southeast" when viewed from a point "way down east". If the hyperbola is referred to its asymptotes as coordinate axes, the equation of the curve becomes $XY = \text{a constant}$.

HYPERBOLAS WITH LARGE ECCENTRICITY

To complete the analysis of hyperbolas and eccentricity, it is in order to consider the case of "squatty" cones. If the apex angle (θ) of the right circular cone is greater than 90° then the angle β is less than 45° . If angle θ were 160° , then β would be 15° .

For hyperbolas created by plane parallel to vertical axis

$$\frac{\sin \alpha}{\sin \beta} = \frac{\sin 90^\circ}{\sin 15^\circ} = \frac{1}{.256} = 3.9 = e$$

For an eccentricity of 10 as shown in Fig. 42, the hyperbola is almost a straight line parallel to Y axis. To achieve this hyperbola, the cone would have to be a very "short" or "squatty" one. The angle β would be $5^\circ 40' 30''$ whose sin is .1.

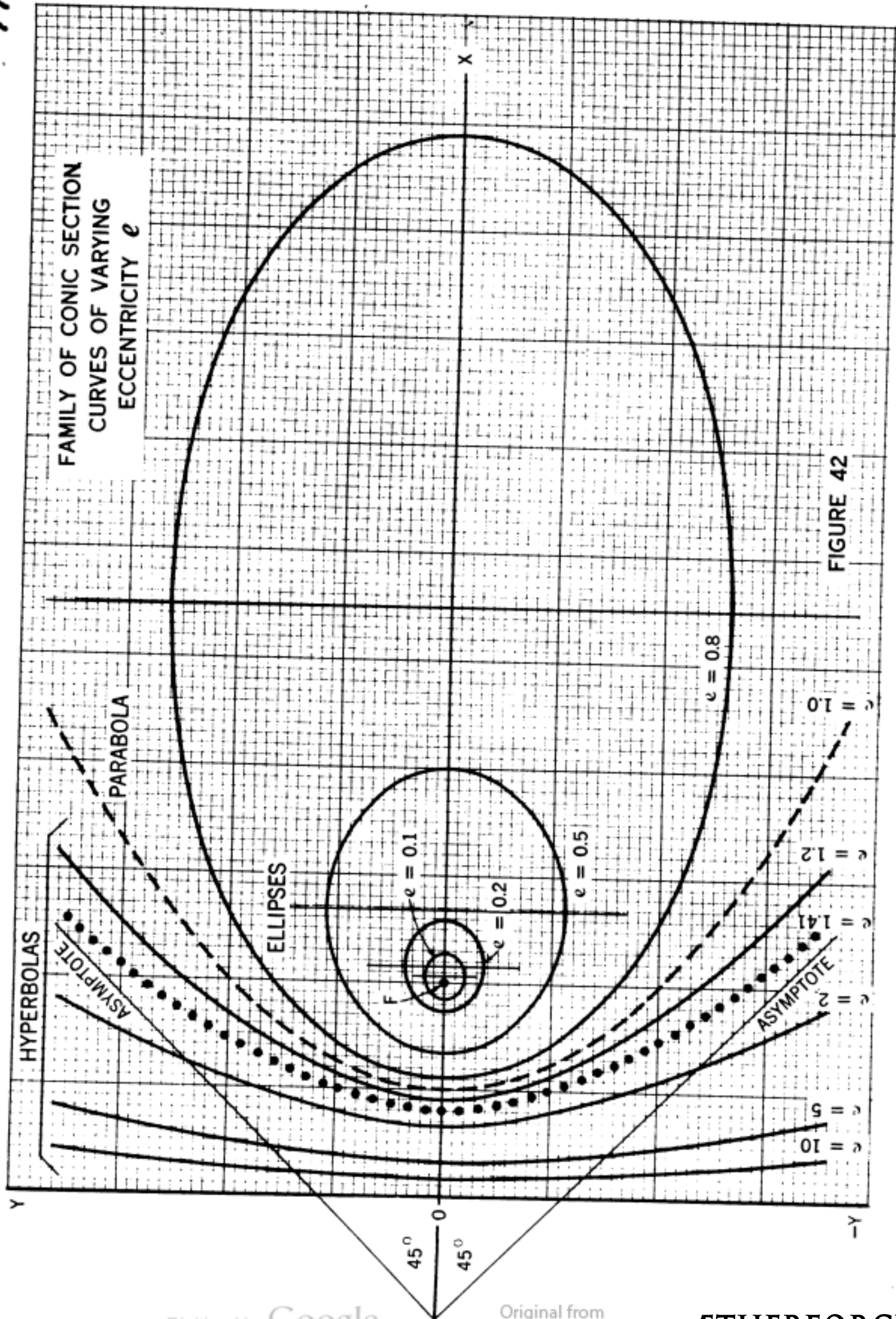
$$\frac{\sin \alpha}{\sin \beta} = \frac{\sin 90^\circ}{\sin 5^\circ 40' 30''} = \frac{1}{.1} = 10 = e$$

The apex angle of the "squatty" cone would be an angle θ of $168^\circ 39'$.

CONCLUSION

The above discussion is based pretty much on two dimensions. To see and sense the conic section curves, one must of course invoke his three dimensional perspective imagination.

The previous photographs of some wooden models (page 8) which were sawed at different angles will greatly assist one's understanding and appreciation of the entire problem. One's three dimensional perspective imagination is provided with an "assist".



SHAPE AND SIZE OF CIRCLES, PARABOLAS, ELLIPSES AND HYPERBOLAS

It is interesting in considering the four conic section curves, that all circles and all parabolas have the same shape while ellipses and hyperbolas vary in shape.

The eccentricity of a circle may be validly thought of as zero (0). The directrix of the circle may be thought of as existing way out to the left of the center of the circle at infinity. The ratio of radius to infinity, $\frac{r}{\infty}$, is always zero whatever the numerical value of the radius. All circles have the same shape although they may be plotted to different scales. There is really but one circle.

This fact results from the fact that there is only one angle of intersecting of a plane with a cone by which a circle is provided. This angle is an angle of 90° between intersecting plane and the altitude line dropped from the apex of cone to the base plane. Of course this altitude line is also 90° to the plane of the base of the cone. The plane of the intersecting circle is parallel to the plane of the base.

All parabolas have an eccentricity of unity (1). By varying the numerical value of " p ", one obtains parabolas of different size but of the same shape. The resulting parabolas will be the same curve drawn to different scales. There is really but one parabola. See previous sketch "A Set of Parabolic Triplets" page 49. They are a set of parabolic arches with displacement on Y axis of varying degree.

This fact results from the fact that there is only one angle of intersection of a plane with a cone by which a parabola is provided. This angle is the angle of the slant surface of the cone with the plane of the base of the cone. The plane of the intersecting parabola is parallel to one slant surface of the cone.

It is not so easy to recognize the exact similarity of shape of varying sized parabolas as it is to recognize the exact similarity of shape of circles of various size.

When one waters the lawn with a hose, the paths of all the various tiny and separate streaks of water are parabolas. The spray is a family of parabolas of the same shape but varying size.

As stated before, it is not easy to regard all these parabolas as of a single shape. In the case of circles the fact of a common shape but different size is readily accepted.

Ellipses and hyperbolas do not have the same shape. The eccentricity of ellipses vary from zero (0) to unity (1). The angle of intersection of plane with cone may vary between the one angle necessary for a parabola (say θ) and the one angle necessary for a circle, 90° .

The eccentricity of hyperbolas varies from unity (1) to very large values. The angle of intersection ϕ of the hyperbolic plane with the vertical height of the cone varies from a maximum value of $(90 - \theta)$ to 0° and then increases its maximum of ϕ as the line AB of intersecting plane is continued to be rotated counterclockwise about point N . See previous Fig. 41.

If line AB makes an angle of zero (0) with base line MN , one gets a circle. If angle BAN varies between zero (0) and θ , one gets an ellipse. If angle BAN is θ , one gets a parabola. If angle BAN varies between θ and $(90 + \theta)$, then one gets an hyperbola.

For an eccentricity of ten (10), the hyperbola almost becomes a straight line parallel to Y axis. It will be pointed out later, that when the eccentricity of the hyperbola is 1.41, it is called an equilateral or rectangular hyperbola.

The accompanying drawing, Figure 42 merits careful inspection and reflective thought.

CHAPTER 8

The Natural Law of Geometric Retrogression: Simple Exponential Equations: Attenuation

In a previous chapter, the sheer facts and the geometric and the trigonometric background of hyperbolic functions were rationalized and made understandable by a simple and straightforward approach.

It is proposed in this chapter to develop and answer to the appropriate question "After all, why talk about and why concern oneself at all with the hyperbolic functions; $\sinh$ and $\cosh$? What are they good for? To express it colloquially: How do they "get into the act"?"

It is hoped that an answer can be arrived at by simple and rational step by step arguments; by an approach involving basic, fundamental and accepted truths and facts which will enable a logical development of a clear path. This path will provide understanding and an explicit, straightforward answer to the question at hand.

It has been said, "only by understanding nature's simple truths can one hope to enlist her as an ally".

PRECISE MEASUREMENT: A BASIC PROBLEM

Engineering and science are basically concerned and are basically built on natural phenomena on one hand and with precise measurement of the various factors involved in the phenomena on the other hand. The engineer and scientist seek to understand thoroughly not only the qualitative but also the quantitative aspects. A knowledge of the sheer qualitative is essential and important but it is not enough to achieve a successful and practical result to provide an integrated system. One must carefully determine relationships by precise quantitative measurement. As said so appropriately by Lord Kelvin, "Until you can measure, you do not know."

ELECTRICAL COMMUNICATION

In electrical communication whether it be provided by sound waves, light waves or electromagnetic waves, the prime objective is to transmit information. In each method and in each system of waves, a small bit of energy, encompassing a band of frequencies, must be transmitted. The wave energy may go its way unguided as far as direction is concerned. This is generally the case in sound, light, and radio waves. In the particular case of electromagnetic waves, wires, cables and hollow wave guides can be used to guide the waves from a given originating point to a particular receiving point, a short or long distance away.

One basic job, aside from the very important matter of switching, is to know not only how the energy is transmitted but also to measure in what degree the energy dies away or "thins out". A good word to specify the "dying away" is attenuation.

It is a common observable fact that the energy of sound waves, light waves and electromagnetic waves, both unguided and guided, decreases with increased distance. As gasoline in the tank of an automobile is used up or attenuates with distance traveled, so does the electrical energy bearing the electrical signal attenuate. The manner of attenuation in the above two cases is quite dissimilar however. These differences will be discussed and compared.

The factors determining electrical energy are voltage, current and time. Time as such does not generally enter the picture, so attention can be focused on voltage and current or their product, namely electrical power.

SINH AND COSH PROVIDE A CONVENIENT TOOL TO CALCULATE LOSSES IN ELECTRICAL CIRCUITS

As a clue to the desired answer, one hopes to get to the original question earlier proposed, it is to be plainly and formally stated at the very beginning that $\sinh$ and $\cosh$ functions are convenient numerical tools to find

power and current and voltage decay in electrical circuits. If one can logically establish quantitative and numerical relationships, by considering and understanding the simple and basic doctrine which nature seems imperiously to decree and rigidly follow in her many phenomena involving attenuation then reference can be readily made to carefully prepared numerical tables of sinh and cosh to provide a quantitative answer. Such quantitative answers must be known in order to design a good working electrical communication system.

Straightforward and simple step by step development of sinh and cosh application will be presented in later pages.

TYPES OF ATTENUATION

The gasoline in the tank of an automobile attenuates directly proportional to distance under ordinary conditions of driving. The gasoline decreases in simple arithmetical progression. However, in most cases of attenuating power in natural processes, the decrease occurs in geometric retrogression fashion.

In many instances of natural phenomena, some of the contributory factors decrease and the decrease occurs in geometric retrogression fashion. Why this is so, why nature follows this method, why nature decrees the doctrine of geometric retrogression, nobody knows. It is her imperious way and it is to be accepted as a basic truth.

The charge on an electrical capacitor decreases with time according to this doctrine when the capacitor is discharged. The electric current in a circuit, containing an inductance with its stored electromagnetic field energy, decreases with time in this geometric retrogression fashion. The amount of disintegrating radium decreases with time according to this method. The voltage, current and power decrease with distance in this fashion along a wire communication circuit.

Attention is particularly directed to the decrease curve of Figures 50, 52, pages 125, 128 and Fig. 67, page 188. The equations of the decrease curves decree geometric retrogression.

Geometric retrogression has many interesting and important features. These will be discussed in later pages.

Suffice it to say now that a clue to the answer sought is that sinh and cosh "get into the act" because of nature's characteristic doctrine of geometric retrogression for attenuation and decay. This doctrine will be thoroughly explored and illustrative arguments will be examined.

In the entire analysis of the general attenuation problem at hand, a common method of nature's attenuation will be presented first.

INVERSE SQUARE LAW

Generally speaking, sound and light energy go out from a source in spherical wave fronts. The energy spreads (like a film of butter) at a particular time over the surface of a sphere. The energy may be thought of a certain thickness. The surface of the sphere is $4\pi r^2$. At a later instant, the energy spreads over the surface of a larger sphere. The film of energy or the energy per square foot on the surface of the larger sphere is thinner. If the larger sphere has twice the radius, the surface area is four times the surface area of the smaller sphere. The amount of energy per unit area of surface is therefore one fourth. If the expanding sphere has at a certain instant, four times the radius, the surface is sixteen times the area of sphere at beginning so the "butter" over the larger sphere is one sixteenth as thick.

So a simple and readily accepted rule about decrease of energy for sound and light states:

The energy per unit area decreases as the square of the distance.

However, light and sound waves soon encounter reflecting surfaces so they do not continue to travel in spherical wave fronts. The simple inverse square law of attenuation in these cases is therefore soon nullified.

Perhaps the case of beamed micro radio waves offers the best illustration of the inverse square law. In this case, a bit of energy goes out from the radio transmitter from a small square mounted horn or square "electrical megaphone" at the source. At some distance away, the energy from the small square area source has spread out over the base of a pyramid. If the distance is 30 miles, the base of the imaginary pyramid is about 1 square mile. The film of energy over this square mile surface is very thin. It is $\frac{1}{400,000}$ as thick as it was at the source of radiation ($.25 \cdot 10^{-5} = 10^{-5.6}$). As will be discussed later this is a 56 decibel loss.

The area of an enlarging square increases as the square of the distance so the energy density decreases as the square of the distance.

In this case too, reflections are likely to nullify strict adherence to the increase square law of attenuation.

USE OF GASOLINE VERSUS DISTANCE

The simplest case of decreasing energy is found in the automobile's use of gasoline. A tank full of gasoline is a supply of energy. The energy in this case decreases directly as the distance traveled as a general truth. Of course at high speed, the automobile's use of gasoline is not at the same rate as at moderate speed but as an overall truth, the energy supply does decrease directly as the distance.

ARITHMETIC RETROGRESSION

If a quantity decreases by the same amount in each of a series of changes, one describes this situation by saying that there has been an arithmetic retrogression.

For each 20 miles of travel, the amount of gasoline in the tank decreases 1 gallon. If on the other hand, some factor increased by the same amount in the series of events then one would have an arithmetic progression. One's salary might be increased \$5. per week for each month of employment for one year. The starting salary was \$25. per week. In one year, the salary would have increased \$5. per week per month. At the end of a year, the salary would be \$85. per week. The increases in salary were a constant amount per month. The salary increases in arithmetic progression. Such a progression would be portrayed graphically by a straight line.

GEOMETRIC PROGRESSION AND RETROGRESSION

In the changes occurring in the factors involved in many natural phenomena, it seems that the changes usually take place in geometric progression or geometric retrogression fashion.

Why this system or method of quantitative change occurs in so many natural phenomena is a question which man cannot answer; it is a basic truth and that's that.

GEOMETRIC PROGRESSION DEFINED AND ILLUSTRATED

Geometrical progression can be considered appropriately in terms of ratios and exponents. Sometimes exponents are called logarithms.

If a given factor (a) varies in geometric progression, it will grow or increase by the same ratio in each successive equal step of the growing process.

In arithmetic progression, the successive quantities can be expressed as follows, where X is the increase or increment and " a " is the starting amount

$$a, a + X, a + 2X, a + 3X, a + 4X, \text{ etc.}$$

In geometric progression, the successive quantities can be expressed as follows, where X is a ratio greater than unity and " a " is the starting amount

$$a, aX, aX^2, aX^3, aX^4, aX^5, \text{ etc.}$$

The X might be designated by " r ". The constant ratio X or " r " can be greater than unity, in which case there is a geometric progression.

If the constant ratio is less than unity, there will be a geometric retrogression.

If the ratio X is assumed to be 2, then the geometric progression series becomes

$$a, 2a, 4a, 8a, 16a, 32a, 64a, 128a, \text{ etc.}$$

MUSICAL SCALE, OCTAVES

Our sense of musical harmony is a good example of a natural phenomena which follows the geometric series doctrine. If one musical note has twice the pitch or frequency of vibration of another note, then our ears are pleased with the harmony of tones which is called an octave.

If one assumes a pitch of just one vibration per second as the "a" in the above illustration, then the frequencies of the successive octaves are 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096.

Actually on the piano, the lowest pitch note is likely to be 32 and the highest pitch note is 4096 vibrations per second.

Furthermore within the octave, the twelve intervening individual steps or half-tones progress in geometric ratio. The ratio in this case is 1.06. If the 512 is multiplied by 1.06 one gets the pitch of the next note, namely 542.7. If this 542.7 is multiplied by 1.06, the pitch of the next note results. The 1.06 is the twelfth root of 2. This means that each white and black key on the piano has a pitch about 6 percent higher than the preceding note. The percentage increase is a constant throughout the growth in frequency of the various notes of the musical scale. The actual amount of pitch increase is not constant; the increment is not constant but the ratio is constant; (1.06).

So if one goes through this process 12 times, namely 1.06 raised to the 12th power (1.06^{12}), the answer is 2 times 512 or 1024.

The succession of the individual notes of a piano provide a pleasing effect to human ears. The increase in their pitch is in geometric fashion or progression.

EXPONENTS, LOGARITHMS

The geometrical series given above was

$$a, 2a, 4a, 8a, 32a, 64a, 128a, 256a.$$

This can be written in a different way; it is a simple and straightforward way.

$$a2^0, a2^1, a2^2, a2^3, a2^4, a2^5, a2^6, a2^7, a2^8.$$

Now we can write down the exponents or logarithms to the base 2, as they are called in logarithmic parlance, which are 0, 1, 2, 3, 4, 5, 6, 7, 8.

The exponents or logarithms progress in arithmetic progression fashion while the numbers of which they are the logarithms progress in geometric progression fashion. Again it is to be said that the so called "base" of the logarithms in this case is 2.

It is to be observed that ($a2^0$) namely "a" times 2 to the zeroth power is "a". Any quantity raised to the zeroth power is 1.

So one can say in words that the pitch of the octaves in a musical scale also vary in geometric progression. Also one can say, the logarithms of the pitch of octaves to the base 2 are directly proportional to the number of the octaves.

If one wants to know the pitch of the 8th octave, assuming the pitch "a" of the first note was 1, then look above the 8 and the answer is 256. In other words, the logarithm of 256 to the base 2 is 8.

A SIMPLE EQUATION

A simple equation can readily and easily be set up to embody the above ideas and facts. The equation will be an exponential equation.

Let Y equal some final term after several multiplications of an initial term by the constant ratio.

Let "a" = magnitude of original term in the geometric series of the growth.

Let r = common ratio (greater than unity) between each pair of the geometric series.

Let N = number of the term in the series.

We can write down in formula fashion, $Y = ar^N$.

This formula states in terse fashion that Y = "a" times a constant ratio "r" which ratio is raised to some power, N .

If r is assumed to be 2 then the beginning or zeroth term of the geometric series is "a", because 2^0 is 1.

If one seeks the 8th term in a series whose zeroth term "a" is 1, then

$$Y = a2^8 = 256a. = 256$$

If the zeroth term "a" were 10 instead of unity (1) then the 4th term would be

$$Y = 10 \cdot 2^4 = 160$$

GEOMETRIC RETROGRESSION

In the above discussion, the constant ratio has been assumed greater than unity. If the common ratio "r" were 1/2 instead of 2, then the simple equation would be

$$Y = a\left(\frac{1}{2}\right)^N$$

If the zeroth or beginning term ("a") were 128, and if N were 4, $Y = \frac{128}{16} = 8$ as the result of 4 intervals of geometric retrogression.

It is easy now to proceed as follows: "r" raised to the N th power = r^N for geometric progression. If ratio is less than unity, namely $\frac{1}{r}$, then $Y = a\left(\frac{1}{r}\right)^N$ for geometric retrogression where "r" itself is greater than unity (1).

The Y , when $N = 1, 2, 3$, etc., will obviously be less than "a" so that one has a geometric retrogression.

It is in accord with simple arithmetical and exponential notation that

$$\left(\frac{1}{3}\right)^2 = \frac{1}{3^2} = 3^{-2}$$

Negative exponents mean that the number is really in denominator

$$Y = a\left(\frac{1}{r}\right)^N = a \frac{1}{r^N} = ar^{-N}$$

In general then in geometric progression, the constant ratio or constant percentage difference means a constant number greater than unity which is to be raised to some positive power; it means a plus exponent or logarithm. If the exponent in the equation is minus, one can expect a geometric retrogression. The simple equation expresses a basic truth in shorthand fashion. The minus exponent need scare no one. It is a code and a signal to indicate a constant percentage loss, a geometric retrogression and attenuation.

COMPOUND INTEREST, GEOMETRIC PROGRESSION

Before continuing on the achievement of our objective (how do sinh and cosh "get into the act"), it will be helpful to explore the geometrical progression and retrogression series doctrine further. Money can be loaned to a bank which pays interest. The interest adds to the principal. The interest can be paid on a compound interest basis; compounded annually, semi-annually, quarterly or even monthly.

Assume one loans \$1. to a bank. The interest is to be so much per cent per annum and is to be compounded yearly. Let the interest rate be "r". The accrued value at the end of any year is readily expressed by the simple exponential formula previously given.

Accrued Value = \$1 (a)ⁿ (where $a = (1 + r)$ and n is the number of years left in bank. If the interest rate is 4 percent then " a " = $1 + .04 = 1.04$. If " n " is zero (0), then " a " or $(1 + r)$ to the zero power is 1 and we have \$1, the original capital. If n is 2 or 5 then

$$\text{Accrued Value} = A. V. = 1 \cdot (1.04)^2 \text{ or } (1.04)^5$$

The accompanying table shows how \$1. would grow of the rate if interest were 10 percent or .1, " a " = $(1 + .1) = 1.1$

$$A.V. = 1(1.1)^n$$

Capital	Percent Interest Compounded Annually	Years	Accrued Value
\$1.	10	0	\$ 1.00
		1	1.10
		2	1.21
		3	1.33
		10	2.59
		11	2.85
		12	3.13
		20	6.73
		25	17.45
		50	117.30
		100	1377.24

The law of geometric growth states that the ratio between two successive terms in the series is a constant. The actual amount of interest received each year increases as the accrued value increases.

$$\$1.33 = 1.1 \text{ times } \$1.21: \quad \text{Interest (3rd yr)} = 22 \text{ cents}$$

$$\$3.13 = 1.1 \text{ times } \$2.85: \quad \text{Interest (11th yr)} = 34 \text{ cents}$$

The law of geometric growth also states that the instantaneous rate at which the accrued value on a particular January 1st increases directly is proportional to the accrued value existent on that same January 1st. This truth cannot be rationalized by inspection or consideration of the above data. If, however, the above data were plotted to constitute a curve with dollars on Y axis and years on X axis, then it could be readily observed that the slope or the $\frac{dY}{dX}$ of the curve at a particular point would gradually increase and be greatest at the \$1377 point.

It will be interesting at this point to comment about gasoline and automobile engines. The engine does not consume gasoline at a rate dependent on the amount of gasoline left in the supply tank. The engine does not use gasoline at a greater rate when the tank is full than when the tank is nearly empty.

The whole compound interest matter as expressed by the equation and illustrated by the above numerical values prompts the observation that the logarithm of the accrued value to the base 1.1 is directly proportional to the number of years. In fact, the number of years is the logarithm. With only \$1 as the assumed beginning capital, this observation is really a mere statement of the basis on which the calculations were made. If the original capital were \$18.75, the observations would be more significant.

One could validly say that 20 is the logarithm of 6.73 to the base 1.1. Also 50 is the logarithm of 117.3 to the base 1.1.

A SPECIAL KIND OF COMPOUND INTEREST

In the discussion and rationalization of $\sinh$ and $\cosh$ per se in a previous section of this treatise, the base of the Napierian system of logarithms was encountered. This so called "natural" system of logarithms has for its base, the unusual number, 2.71828. This number is conventionally designated as e , which is the Greek letter epsilon. Oftentimes, just plain "e" of the English alphabet is used. The numerical values of $\sinh$ and $\cosh$ are provided by relations to the numerical values of the exponentials, e^x and e^{-x} . Here is a clue as to why $\sinh$ and $\cosh$ "get into the act". The clue will be thoroughly explored later.

As has been previously indicated there are no intrinsic definitive or interpretable geometric relationships between e^x , e^{-x} and $\cosh x$, $\sinh$. The relationships may be said to be purely mathematical and numerical.

To say that $\cosh X$ is defined by its numerical equivalency to $\frac{e^x + e^{-x}}{2}$ is neither valid nor tenable in the usual meaning of the word defined.

ORIGIN OF e , THE BASE OF NAPIERIAN LOGARITHMS

The binominal theorem provides an "assist" in arriving at the numerical value of e . The algebraic expression $\left(1 + \frac{1}{n}\right)^n$ is expanded by the binominal theorem technique. If " n " is infinite in value, then 2.71828 becomes the numerical value of the algebraic expression and is designated by e or e . Its value to more decimal places is 2.71828 18284 59045.

In considering the geometric progression and geometric attenuation in general, the compound interest and compound spending doctrine was invoked to provide background and understanding.

COMPOUND INTEREST OF A SPECIAL KIND

The determination of the numerical value of e can be found by using an interesting analogy to compound interest of a rather unconventional type.

Suppose a banker said, "You put in \$1. in our bank and during the first year we will pay you 100 percent interest per year." The accrued value of the \$1. at the end of the first year will, of course, be \$2.

The general compound interest formula in this case would be

$$\left(1 + \frac{1}{1}\right)^1 = 2$$

Then the banker continues, "But, if you have your \$1. in the bank for 2 years we will pay you 50 percent interest compounded for the two years." The saving bank depositor might think this a fine deal and readily subscribe to the new kind of savings account. However, it would not be a good deal, for the banker has more to say.

For two years, the general formula provides the following amount of accrued to value.

$$\left(1 + \frac{1}{2}\right)^2 = 2.25$$

Then the banker continues, "But if you let the money stay in the bank for 3 years we will pay you 33 $\frac{1}{3}$ percent interest compounded for the 3 years. If you let us keep the money for 4 years we will pay you 25 percent interest for 4 years and so on and on for many years.

The formulas would be $\left(1 + \frac{1}{3}\right)^3$, $\left(1 + \frac{1}{4}\right)^4$, $\left(1 + \frac{1}{11}\right)^{11}$, $\left(1 + \frac{1}{30}\right)^{30}$, $\left(1 + \frac{1}{n}\right)^n$.

The accrued value on this kind of a compound interest basis would after many many years amount to only \$2.71828. The deal for the first year was fine providing \$2. at the end of the first year but for many many years more, the additional accrued value would be only 72 cents.

The following table provides interesting inspection

Original Capital	Interest (%)	Years	Accrued Value
1	100	1	2.00
	50	2	2.25
	33 $\frac{1}{3}$	3	2.37
	25	4	2.44
	20	5	2.488
	10	10	2.5937
	9.09	11	2.606
	8.33	12	2.613
	5	20	2.648
	5.33	30	2.673
	2.	50	2.691
	1	100	2.70481
	0.1	1000	2.71692
	0.01	10000	2.71815
	0.001	10000	2.71827
	0	Infinite	2.71828

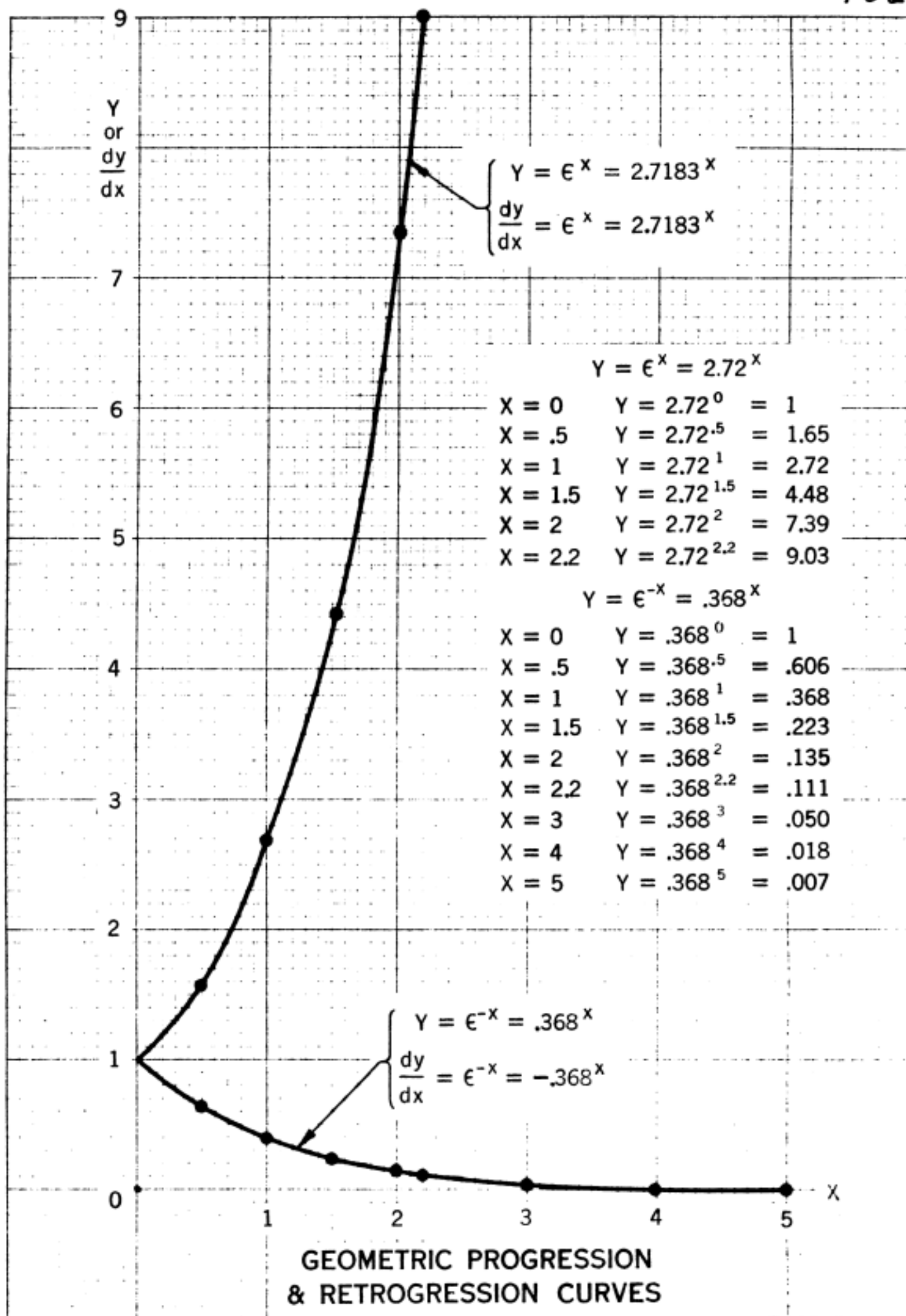


FIGURE 43

The last figure results from a seemingly silly proposition. If the \$1. is left in the bank after a very, very great number of years, practically no further interest at all will be paid but nevertheless the accrued value will be \$2.72.

In passing, it is in order to note that the numerical value of e can be calculated by the series expansion doctrine

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots$$

OTHER FORMULAS FOR COMPOUND INTEREST

As pointed out, the simple equation for compound interest at 10 percent is $Y = 1 \cdot (1.1)^N$

This equation says that N is the logarithm of Y to the base 1.1. It would be easily evaluated if one had a table of logs for numbers to the base 1.1. Since there is no such table available, one can translate 1.1 to the base 10 or common logarithms. Since $1.1 = 10^{.04148}$, we can write the equation $Y = 1 \cdot 10^{.04148 N}$.

The evaluation of the equation for any value of N is then easy according to simple logarithmic rules. Furthermore, since there are tables of Napierian logarithms with e (2.7181) as a base, we can ask, to what power must e be raised to give 1.1. This power is .0954.

So the equation can now be written in another form.

$$Y = 1 \cdot e^{.0954 N}$$

In conclusion then, we can have a compound interest equation founded on three logarithmic bases to specify geometric progression or natural growth. There is no preference among the equations except $Y = 1 \cdot e^{.0954 N}$ can be more easily differentiated as per calculus.

There are tables for the latter two logarithmic bases, so they permit easy numerical calculation for Y for large values of N .

COMPOUND SPENDING: GEOMETRIC RETROGRESSION: ATTENUATION

Assume one started off on a modest vacation trip with \$100 and planned to spend an amount of money each day according to the geometric retrogression policy. Assume one decided to spend each day 10 per cent of the amount at hand at beginning of each day. The financial problem for the first few days would be readily calculated by simple arithmetic and the results would be as follows.

Day	Money on Hand	Amount Spent (10%)	Amount Left (90%)
1st	\$100.00	\$10.00	\$90.00
2nd	90.00	9.00	81.00
3rd	81.00	8.10	72.90
4th	72.90	7.29	65.61
5th	65.61	6.55	59.05
6th	59.05	5.91	53.14

For geometric retrogression herewith involved the formula readily becomes

$$\text{Sum left} = 100 \cdot (.9)^N$$

So being interested at the beginning of the vacation, in amounts he will have available to spend in the latter part of the vacation, the man figures matters out ahead of time according to the above attenuation formula and comes up with the following data.

Day	Money on Hand	Amount Spent (10%)	Amount Left (90%)
15th	\$22.27	\$2.22	\$20.05
20th	13.50	1.35	12.15
21st	12.15	1.21	10.94
35th	2.78	.28	2.50
40th	1.64	.16	1.48
50th	.57	.05	.52
60th	.20	.02	.18

If there were no special reserve, it would be time to telegraph home for more money.
 The use of the formula for the latter days provides some interesting thinking.
 The formula for the 20th day would be

$$\text{Sum left} = 100 \cdot (.9)^{20}$$

The shortest way to get the answer is to use logarithms, either common logarithms to the base 10 or Napierian logarithms to the base e or 2.71828.

So one writes the equation in three different ways.

$$\begin{aligned} Y &= 100 \cdot (.9)^N \\ &= 100 \cdot 10^{-.0457 N}; \quad (10^{-.0457} = .9) \\ &= 100 \cdot e^{-.1054 N}; \quad (e^{-.1054} = .9) \end{aligned}$$

SOME MORE ABOUT MUSICAL SCALE: A GEOMETRID PROGRESSION

Instead of considering only the octaves of musical tones whose frequency vary in geometric progression by a constant ratio of 2, let us consider the semi-tones on the piano. There are as previously stated, 12 intervals or 12 semi-tones in the octave. Therefore each semi-tone (increasing in geometric progression), is the twelfth root of 2 in value, namely 1.05926. This is saying again that each semi-tone on the piano is about 6 percent higher in frequency than its predecessor.

We can write the exponential equation for the pitch of any key on the piano starting from a lower "C" of 32 vibrations per second.

$$\text{Pitch} = 32 \cdot (1.05926)^N$$

The base of the logarithms in this case is 1.05926 and the log is N . But since no log tables to this base are generally available, we can better use either common logarithms or Napierian logs; whichever suits our fancy.

We first ask; 10 raised to what power gives 1.05926 or what is log of 1.05926 to base 10. Of course it is one twelfth of log 2 to base 10, namely $.30103 \div 12$ or $.02508$.

So the equation can now be written in terms of logs to the base 10.

Pitch of any key which is N semi-tones above 32 vibrations per second = P . $P = 32 \cdot 10^{.02508 N}$

We might like to write the equation in terms of Napierian base e (2.7183). Logs to base e are 2.30258 times logs to base 10. The answer to our question therefore is $e^{.0576} = 1.05926$ since $2.30258 \text{ times } .02508 = .0576$.

Of course from a table of logs to base "e", the log of 1.05926 to base "e" can be found as $.0576$. In other words, an e^x table could be consulted.

So the equation can be now written in a third form.

$$\text{Pitch} = 32 \cdot e^{.0576 N}$$

To repeat; the three exponential equations are

$$P = 32 \cdot 1.0596^N$$

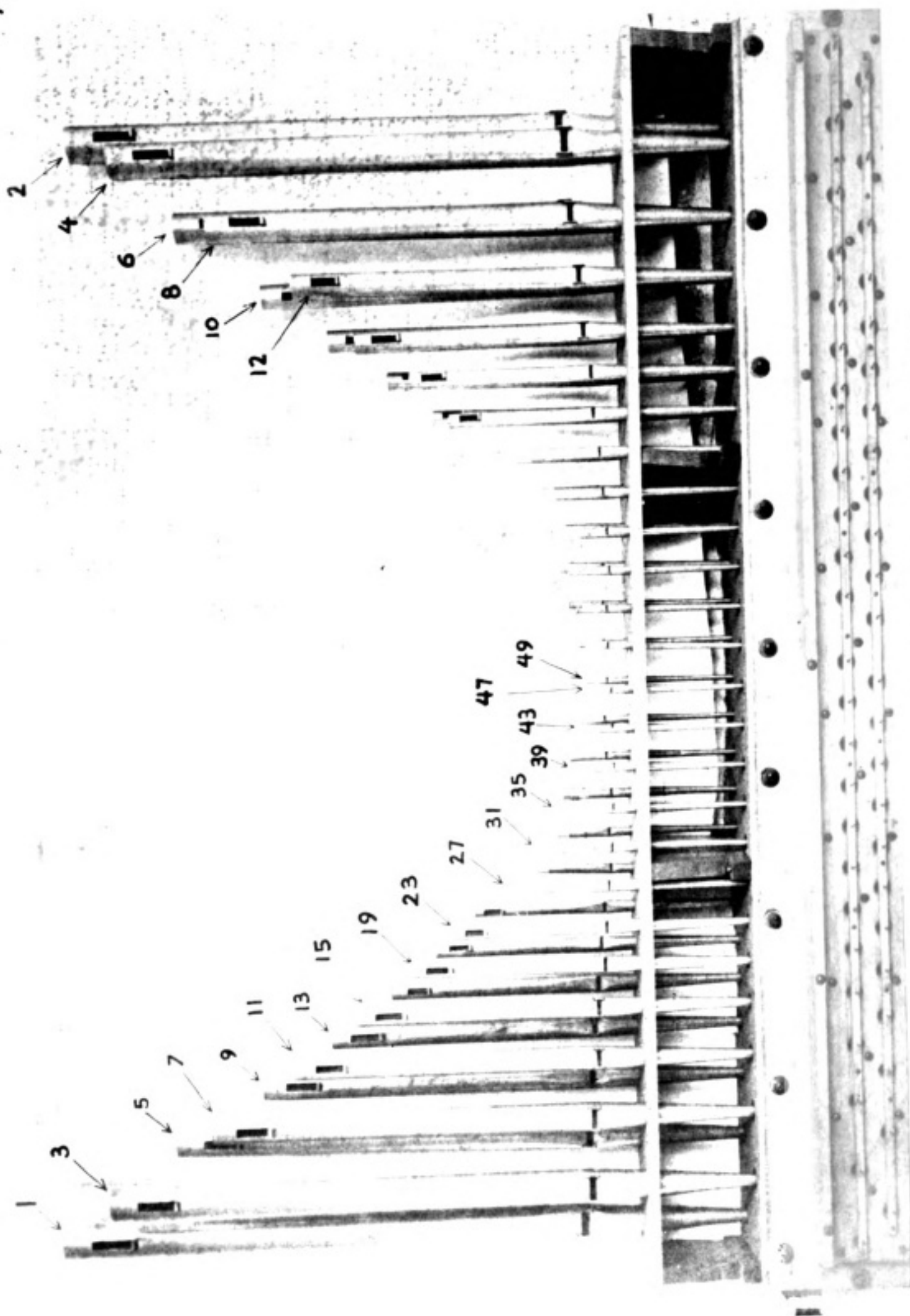
$$P = 32 \cdot 10^{.02508 N}$$

$$P = 32 \cdot e^{.0576 N}$$

The above equations involve three logarithmic bases. The two conventional logarithmic bases are 10 and 2.7182. Extensive tables of numbers to these two logarithmic bases have been prepared with fine accuracy. It is interesting to observe that no letter has been conventionally designated to specify the logarithmic base 10. The 2.7182 base is conventionally designated to be e ; epsilon or just plain e .

Interrelations of Logarithmic Bases

The following simple expression provides the relations of the log of a given number to one base to the log of the number to another base.



A FOUR OCTAVE ARRAY OF OPEN ORGAN PIPES
256 (NO. 1) TO 4096 (NO. 49) VIBRATIONS PER SECOND
FIGURE 44

$$\log_a X = (\log_b X) \log_a b$$

For the two conventional logarithmic bases, the relations are

$$\log_e X = \log_e 10 \log_{10} X = 2.3026 \log_{10} X$$

$$\log_{10} X = \log_{10} e \log_e X = .4343 \log_e X$$

GEOMETRIC PROGRESSION EQUATIONS

The first exponential equation of the foregoing three is the really important and significant equation of the three. This first equation expresses the simple and basic truth of a geometric progression. It is easily and readily interpretable and understandable in its embodiment of the geometric progression doctrine.

The second and third equations are not equations which in themselves embody the basic truth at hand. They are equations which give instructions how to get a numerical answer to the first equation. They are equations which suggest and provide the easy method of logarithms to get an answer other than the laborious process of arithmetical multiplication. The logarithmic bases "e" and "10" do not contribute in any way to the understanding or mathematical expression of the basic concept of geometric progression.

One need not wonder nor be concerned how they (the "e" and "10") get into the act. They are only mathematical tools. The first equation is the one that gives the real story and the mathematical expression of a simple and understandable principle.

All three exponential equations express a natural growth or natural geometric progression of exactly the same degree. The last equation does not express natural growth any better than the first two. The logarithmic base of "e" is sometimes more convenient but it does not provide growth equations that are more natural than do the bases 10, 1.05926 or any other base. The numerical value of e is 2.71828, so it is quite a sizable constant ratio as of course 10 is too. Such large ratios of these exact values do not generally prevail in natural phenomena so exponents to the two bases are small such as the .02508 and .0576.

LENGTH OF ORGAN PIPES—AN EXAMPLE OF GEOMETRIC PROGRESSION

The accompanying photograph Fig. 44 of an array of organ pipes nicely illustrates a geometric progression and exponential change. The interesting fact that the frequencies of the musical scale increase in geometric progression fashion has already been discussed. The length of organ pipes is intimately related to the frequency of the emitted tone. The column of air in the pipe resonates with the frequency produced at the lip of the pipe. The length of the resonating column of air is in general one fourth or one half of the emitted sound wave length in air depending upon the fact of the pipe being a closed (stopped) pipe or an open pipe respectively. This one fourth and one half wave length concept might be said to be a theoretical relation but practically the relation is valid to a close degree of approximation. For a pitch of 256 vibrations per second, the sound wave length (closed pipe) would be 4.26 feet, assuming the velocity of sound at 32° Fahrenheit is 1090.6 feet per second. An open organ pipe emitting 256 vibrations per second would be 2.13 feet long.

As the pitch of the sound increases, the length of the pipe decreases. Since the pitch of the successive musical notes increases by equal ratio amounts so the lengths of the pipes decrease by equal ratio amounts. Conversely as the pitch decreases in geometric retrogression fashion, so the lengths of the pipes increase in geometric progression fashion.

The lowest note emitted by the array of organ pipes shown in photographs is 256 vibrations per second. The length of this pipe at the left (No. 1) is 24 inches. There are forty nine pipes for the four octaves with forty eight constant ratios of 1.059 involved. The shortest pipe is in center of back tier (No. 49) and has a pitch of 4096 vibrations per second and a length of 1.6 inches.

Due to the symmetrical arrangement of the pipes in an actual organ, one cannot observe the exponential curve for any one of the four octaves. As noted in photograph the pipes of the first octave are numbered 1-13. These thirteen pipes are so placed to fit in an overall attractive and balanced "layout".

Musical Note	No.	Constant Ratio 1.059	Pitch Vib. Per Sec.	Length Inches	Constant Ratio 1.059
Do C.....	1		256.0	24 (2.000)	
C#.....	2	1.059	272.1	22.4 (1.879)	
Re D.....	3	1.059 ² (1.096)	287.1	21.2 (1.774)	
D#.....	4	1.059 ³ (1.193)	304.1	20.1 (1.675)	
Mi E.....	5	1.059 ⁴ (1.258)	322.0	18.9 (1.582)	
Fa F.....	6	1.059 ⁵ (1.332)	341.0	17.5 (1.494)	
F#.....	7	1.059 ⁶ (1.411)	361.0	16.7 (1.411)	
Sol G.....	8	1.059 ⁷ (1.494)	382.4	15.9 (1.332)	
G#.....	9	1.059 ⁸ (1.582)	404.8	14.9 (1.258)	
La A.....	10	1.059 ⁹ (1.675)	428.9	14.1 (1.193)	
A#.....	11	1.059 ¹⁰ (1.774)	454.3	13.2 (1.096)	
Ti B.....	12	1.059 ¹¹ (1.879)	480.1	12.6 (1.059)	
Do C.....	13	1.059 ¹² (2.000)	512.0	12.0	

As stated before, the successive pipes of a particular octave do not show graphically an exponential curve. However if one considers the twelve pipes in the front row of pipes at the left, the rear row at the left, the front and rear rows at the right, one does observe an exponential curve. The front left row are numbered on the photograph and are three notes of each of the four different octaves. The constant ratio between the individual lengths and pitches of these pipes involving four basic intervals is 1.059⁴ or 1.258. So one has four series of pipes and four exponential curves, with a constant ratio of increment of 1.258.

The equation of each of the four curves is

$$L = L_0 1.258^N = L_0 10^{.0996N} = L_0 e^{.2296N}$$

Musical Note	No.	Constant Ratio 1.258 (1.059) ⁴	Pitch Vib. Per Sec.	Length Inches	Constant Ratio 1.258 (1.059) ⁴
D ₁	3		287.1	21.1	
F ₁ #.....	7	1.258	361.0	16.7	
A ₁ #.....	11	1.258 ² (1.582)	454.3	13.2	
D ₂	15	1.258 ³ (2.000)	574.2	10.4	
F ₂ #.....	19	1.258 ⁴ (2.502)	722.0	8.6	
A ₂ #.....	23	1.258 ⁵ (3.164)	908.6	6.8	
D ₃	27	1.258 ⁶ (4.000)	1184.4	5.3	
F ₃ #.....	31	1.258 ⁷ (5.000)	1444.0	4.2	
A ₃ #.....	35	1.258 ⁸ (6.328)	1817.2	3.3	
D ₄	39	1.259 ⁹ (8.000)	2296.8	2.6	
F ₄ #.....	43	1.258 ¹⁰ (10.07)	2888.0	2.1	
A ₄	47	1.258 ¹¹ (12.66)	3634.4	1.8	

THE EXPONENTIAL HORN

The shape and contour of a so called exponential horn is a fine example of a geometric progression curve and doctrine. The radius of the horn increases in a constant ratio as a function of the equal increments of length of the horn. See Figure 45.

The equation of the curve of the horn can be readily written in three forms.

$R = R_0 r^L = R_0 e^{CL} = R_0 10^{KL}$ where R is the radius at any distance along the length, r is the constant ratio, R_0 is the radius of the throat and L is the length along horn axis. The C and K are logarithmic increment constants.

It is in order to consider the cases of two horns. In the first case, one desires a horn whose throat is to be one inch in radius, whose mouth is sixteen inches in diameter (8 inches in radius) and whose total length is

STRAIGHT LINE HORN VERSUS EXPONENTIAL HORN

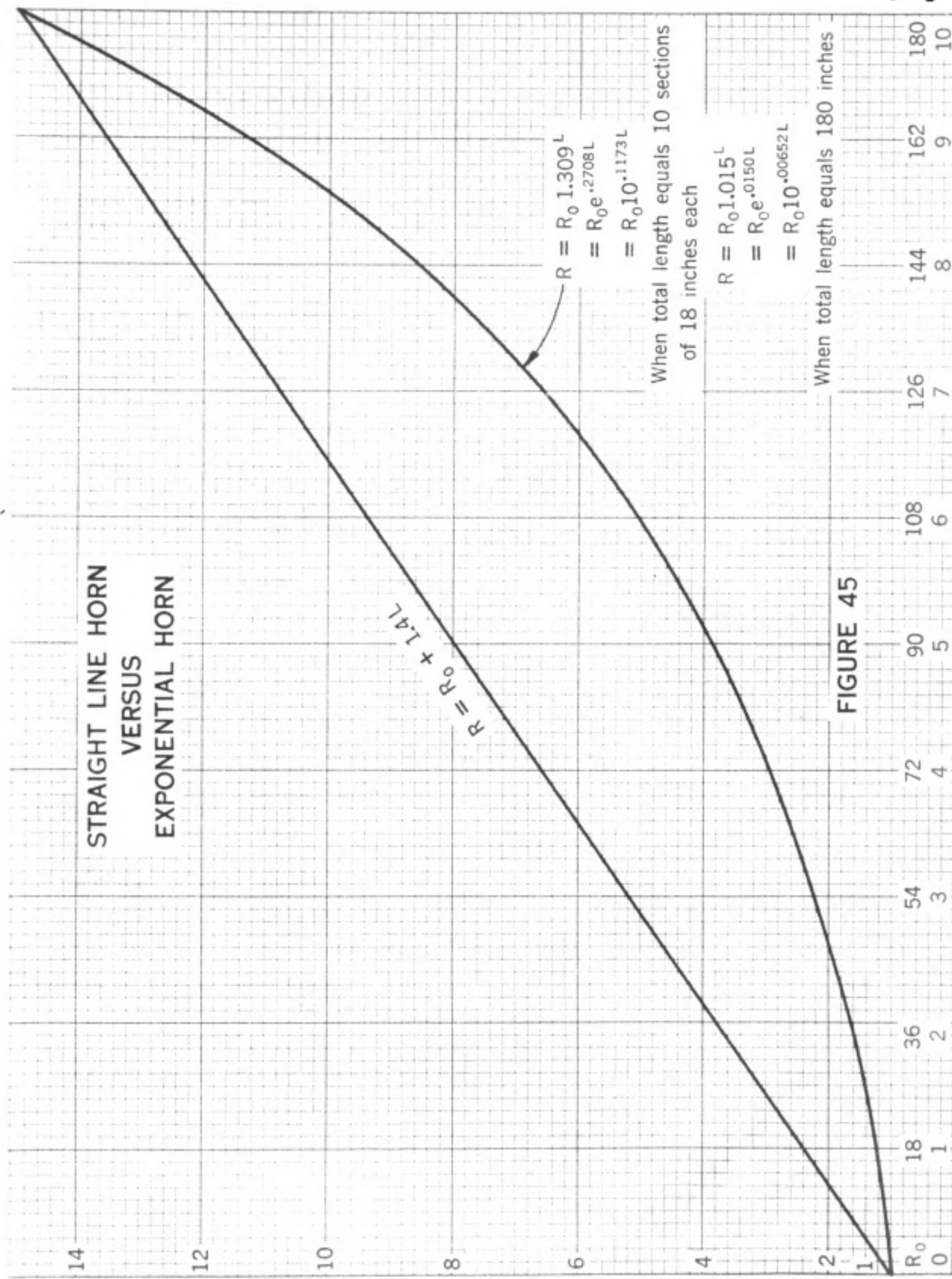


FIGURE 45

60 inches (5 feet). The equation in this case as per above general equation would be

$$8 = 1 \cdot r^{60} = 1 \cdot e^{C \cdot 60} = 1 \cdot 10^{K \cdot 60}$$

It is desired by find the proper "r", C and K for this particular horn.

Writing the above equation in terms of logarithms to the base "e" for convenience

$$\log_e R = \log_e R_0 + \log_e e^{C \cdot L}$$

$$\log_e 8 = \log_e 1 + 60C$$

The log of a product is the sum of the logs and $\log_e e^{C \cdot 60}$ is $60 \log_e e^C$ and $\log_e e^C$ is C itself.

Therefore $2.0795 = 0 + 60C$. $C = .0346$ and $K = .4343$ times $.0346 = .0150$

Therefore the "r" which it is desired to know and which equals e^C and 10^K is equal to $e^{.0346} = 10^{.0150}$. As per each log table, the "r" desired = 1.035

The equations can now be written

$$R = 1 \cdot 1.035^{60} = 1 \cdot e^{.0346 \cdot 60} = 1 \cdot 10^{.0150 \cdot 60} = 8 \text{ inches}$$

This means that the radius of the horn increases by the constant ratio 1.035 for each inch of length. The radius at 10 inches would be $R = 1 \cdot 1.035^{10} = 1 \cdot e^{.0346 \cdot 10} = 1 \cdot e^{.346} = 1.413 \text{ inches} = 1 \cdot 10^{.0150 \cdot 10} = 1 \cdot 10^{.150}$. For each 10 inches of length, the radius increases 1.413 times.

For the total length of 60 inches or six sections of 10 inches each the radius of the mouth of the horn would be

$$R = 1 \cdot 1.413^6 = 1 \cdot e^{.346 \cdot 6} = 1 \cdot e^{2.076} = 1 \cdot 10^{.150 \cdot 6} = 1 \cdot 10^{.900} = 8 \text{ inches}$$

This horn would be a so called "morning glory" or exponential loud speaker horn which was quite the vogue yesteryear.

As another example, consider a large or so called "bull" exponential horn. Assume that it is desired that the throat be 2 inches in diameter (1 inch radius), that the mouth of the horn be 30 inches in diameter and that the total length be 180 inches (15 feet).

The basic and simple equation would be

$$R = R_0 \cdot r^{180} = R_0 \cdot e^{180 \cdot C} = 15 \text{ inches}$$

As before for purposes of accurate design we want to find "r". Writing the equation in terms of logarithms to base "e" for convenience as before

$$\log_e 15 = \log_e 1 + 180C$$

As per log tables to base "e" available

$$2.708 = 0 + 180C$$

$$C = .0150 \text{ and } K = .0150 \text{ times } .4343 = .00652$$

Therefore the "r", which we are seeking is

$$e^{.0150} = 10^{.00652} = 1.015$$

This means that for each inch of length, the ratio of a radius to the previous radius is 1.015.

The general equation then takes on definite quantitative values.

$$R = 1 \cdot 1.015^{180} = 1 \cdot e^{.0150 \cdot 180} = 1 \cdot e^{2.708} = 15 \text{ inches}$$

$$= 1 \cdot 10^{.00652 \cdot 180} = 1 \cdot 10^{1.173} = 15 \text{ inches}$$

As a matter of further analysis, consider the horn in 10 sections of 18 inches each. For an L of 18 inches

$$\begin{aligned} R &= 1.1015^{18} = 1 \cdot e^{.0150 \cdot 18} = 1 \cdot e^{.2708} = 1.1.309 \\ &= 1 \cdot 10^{.00452 \cdot 18} = 1 \cdot 10^{.1173} = 1.1.309 \end{aligned}$$

This means that for every 18 inch section of length, the ratio of radius increase is a constant of 1.309. So for the entire horn of 10 increments of 18 inches each the equation can be readily written

$$\begin{aligned} R &= 1.1.309^{10} = 1 \cdot e^{.2708 \cdot 10} = 1 \cdot e^{2.708} = 15 \text{ inches} \\ &= 1 \cdot 10^{.1173 \cdot 10} = 1 \cdot 10^{1.173} = 15 \text{ inches} \end{aligned}$$

The curve of the shape or contour of this "bull" horn would have a 30.9% constant ratio of increase of radius for each 18 inches of length. A horn with straight line for the boundary contour would have an arithmetic increase of 1.4 inches per 18 inches section of length.

According to accompanying curve one could regard the straight line horn and the exponential horn to be 10 units long or 180 inches long according to scale of distances on horizontal center line of horn. The "r" for each of the ten units of length is 1.309 and for each inch of length the "r" is 1.015.

The first horn herewith considered might be thought of as a 60 inch horn with "r" = 1.035 or a horn of 10 unit lengths with an "r" of 1.413.

Today big exponential horns are used for the public address system in stadiums and similar places.

In the above discussion, the contour of only one side of the exponential horn has been considered. The reader can have a good impression of an entire horn by taking a look at the exponential curves of Fig. 46, page 116. By looking from the east as it were in this figure, the shape of the entire expanding horn is well illustrated.

EXPONENTIAL AREA

The cross sectional area of an exponential horn also follows the exponential law.

If the radius of a "morning glory" circular cross sectional horn has a constant ratio rate of increase per unit of length of 1.015, the area ratio rate of increase would be the square of 1.015 or 1.030. The logarithmic constant for the area would be twice that of the radius.

The area equations would be

$$\begin{aligned} A &= A_0 1.030 = A_0 e^{2 \cdot .0150 L} = A_0 e^{.030 L} \\ &= A_0 10^{2 \cdot .00452 L} = A_0 10^{.012 L} \end{aligned}$$

RADIUM AND RADIOACTIVITY

An excellent example of the existence of geometric retrogression and attenuation as the imperious quantitative law in a natural phenomenon is found in the case of the life of radium and other radioactive material. These materials disintegrate into other forms; they are constantly decreasing with time in amount of their very selves into other forms of matter. A milligram of radium is a milligram of radium for only a short time. Later there is less than a milligram and wee bit of something else, a wee bit of some other matter. The milligram of radium lives only a little while as a milligram of radium. It has been found by careful and precise measurement that in 1580 years, a milligram of radium has decreased to a half of a milligram of radium and a half of a milligram of something else. The period of half life of radium, as the phrase is used, is 1580 years.

The imperious law of geometric retrogression decrees that in another 1580 years, half of what was left at the end of the first 1580 years will have disintegrated. Of course one half of one half is one fourth. It took 1580 years for the original amount to disintegrate to one half but it also takes 1580 years for the original amount to disintegrate to one fourth ($\frac{1}{2}$ of $\frac{1}{2}$). The radium has lived 3160 years but one quarter of its original amount remains. After the next 1580 years $\frac{1}{2}$ of $\frac{1}{4}$, will be left. In other words, the percentage decrease each 1580 years is a constant namely 50 per cent or a ratio of $\frac{1}{2}$. It is a matter of $\frac{1}{2}$ of $\frac{1}{2}$ of $\frac{1}{2}$ and so on. Theoretically the original milligram of radium will never vanish entirely. The equation can be written

$$M = 1 \text{ milligram} \left(\frac{1}{2} \right)^{100} = 1 \cdot \frac{1}{2^{100}} = 1 \cdot 2^{-100}$$

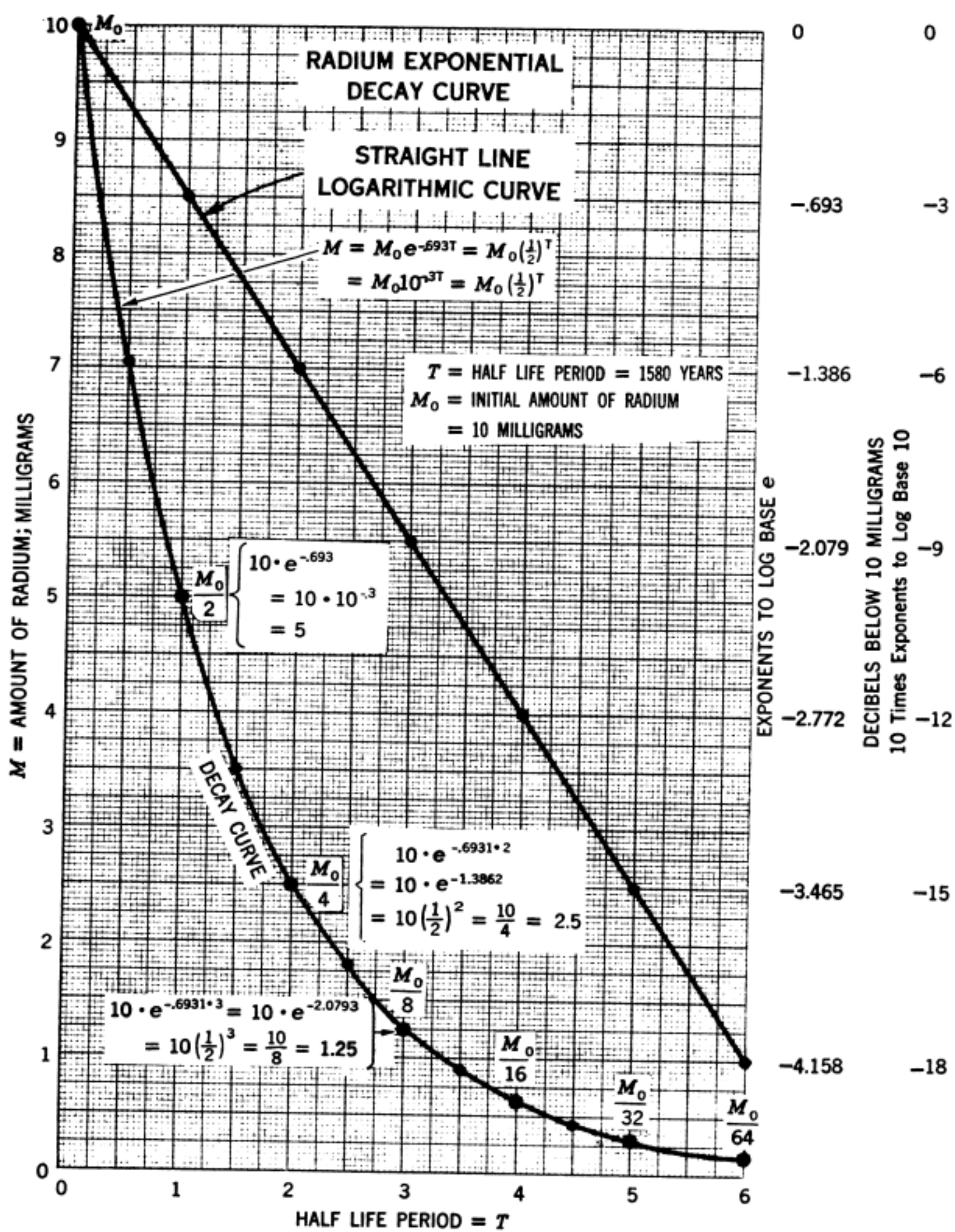


FIGURE 45A

This 100 is the number of 1580 year periods that one might choose to talk about. The period of "half life" of radium is 1580 but it has many of these "half life's".

Uranium 235 has a "half life" of 7.07×10^8 years.

Thorium has a period of "half life" of 2.2×10^{10} years.

Actinium has a "half life" of .002 second.

Polonium 212 has a "half life" of 3×10^{-7} seconds.

It would not be easy, except by a lot of simple arithmetic ($\frac{1}{2}$ multiplied by itself 100 times) to evaluate the expression for amount of radium left after 100 "half life" periods of 1580 years each.

$$M = 1 \text{ milligram} \left(\frac{1}{2}\right)^{100} = 1 \cdot 2^{-100} = 1 \cdot \frac{1}{2^{100}}$$

is a typical example of geometric retrogression or attenuation characteristic of so many decays in nature.

One would need to resort to logarithms to base 10 to conveniently get an answer to the above problem, since there are no tables to the base 2 available. One writes the equation in terms of logarithms to the base 10. The log of 2 to base 10 is .30103. Therefore,

$$M = 1 \cdot \left(\frac{1}{2}\right)^{100} = 1 \cdot \left(\frac{1}{10^{.30103}}\right)^{100} = 1 \cdot 10^{-.30103 \cdot 100}$$

The value for M would of course be very, very small. One milligram of radium whose half life is 1580 years would attenuate to a very, very, very small amount in 158,000 years (100 "half lives"); $(\frac{1}{2})^{100}$ is very small.

EQUATION OF RADIUM EXPRESSED IN NAPIERIAN AND BRIGGSIAN LOGS

If the only log table available were that of Napierian logarithms, then we could ask to what power must e be raised to get 2. It is .6931: $e^{.6931} = 2$. Then we can write above equation in different form.

$$\begin{aligned} \text{Amount of radium} = M &= 1 \left(\frac{1}{e^{.6931}}\right)^{100} = 1 \cdot e^{-.6931 \cdot 100} \\ &= 1 \cdot 10^{-.30103 \cdot 100} \\ &= 1 \cdot \left(\frac{1}{2}\right)^{100} \end{aligned}$$

RADIUM DECAY EXPRESSED IN DECIBELS

As stated above, the amount of radium remaining from an initial amount of one milligram would be very, very small after 100 half lives of 1580 years each. If one did wish to calculate the very small amount, the sheer arithmetic would be terrific. The equation $M = 1 \cdot (\frac{1}{2})^{100}$ would call for a lot of multiplying as such.

If one chose to use the above logarithmic exponential equations, one would encounter very small numbers whose logarithms are not easy to manipulate. The sheer arithmetic would again be quite sizable. However, one can take advantage of the fact of one half decay in 1580 years and adopt a simple alternative. One can write $\frac{1}{2}$ as $\frac{1}{10^{.3}}$ or $10^{-.3}$, (a .3 bel or 3 decibel loss).

In ten half life periods, $\left(\frac{1}{10^{.3}}\right)^{10}$ would be involved. The numerical value of this expression is .001 which is $\frac{1}{10^3}$ or 10^{-3} (a 3 bel or 30 decibel loss).

One can use the language of bels and decibels. The bel is the name given to the exponent of 10. One bel = 10 decibels. So in 10 half lives, the 1 milligram of radium would be .001 of a milligram. The language of decibels says, "There has been a loss of 3 bels or 30 decibels" to indicate the loss to .001. The 30 decibels is a term to mean a decay to .001.

In the second span of 10 half lives, there would be another decay of .001 or 30 decibels. The amount left would be another decay of .001 or 30 decibels. The amount left would be .001 times .001 or .000001 milligrams;

10^{-6} . There would be a 60 decibels loss. In a third span of 10 half lives involving a total of 47,400 years (30 times 1580), the amount left would be "down" 90 decibels; 10^{-9} or .000000001 milligrams.

FINAL EQUATION FOR RADIUM DECAY

If radium decreases by one half in 1580 years, one might appropriately ask how much does it decrease in one year as it decays in geometric retrogression fashion. To answer this question, we ask what is the 1580 root of 2. Some arithmetic and the use of log tables provides the answer namely that the 1580 root of 2 is 1.00042; $^{1580}\sqrt{2} = 1.00042$. This means in turn that

$$1.00042^{1580} = 2.$$

So we can write the radium decay equation in the basic simple form

$$\begin{aligned} M &= Mo \left(\frac{1}{1.00042} \right)^N \\ &= Mo .99957^N \end{aligned}$$

In this equation N is the number of years. The basic and simple truth is that 1 milligram of radium would become .99957 of a milligram in one (1) year. After two years, there would be .99957 times .99957 of a milligram; .99914

$$M = Mo .99957^{1580} = \frac{Mo}{2} \text{ in 1580 years.}$$

Finally then we can write the decay of equations for radium in three different ways. In all of the following equations N is the number of years

$$\begin{aligned} M &= Mo (.99957)^N \\ &= Mo 10^{-.000190N} \quad (10^{-.000190} = .99957) \\ &= Mo e^{-.000428N} \quad (e^{-.000428} = .99957) \end{aligned}$$

The first equation expresses the simple and basic truth of geometric retrogression. The second and third equations suggest that the easy way to get a numerical answer is to use logarithms. It is to be stated here as it was stated in the case of geometric progression; one need have no fear of "e" and "10". They make no contribution to the physical interpretation and understanding of natural decay. They only "get into the act" because they are good tools for calculation and getting a numerical answer.

The Naperian base "e" is not a particularly endowed and esoteric natural number. It is rather an artificial number. It is a sheer numeric. It has no intrinsic quality nor character. It has no physical significance which provides insight in the modus operandi of a phenomenon. The mere sight of "e" need scare no one.

AN ASSOCIATE IDEA OF GEOMETRIC RETROGRESSION OR ATTENUATION

In geometric attenuation, the percentage decrease of some factor per unit time or per unit distance is a constant. This truth is a fundamental and simple truth.

A second fundamental truth is that the instantaneous rate of change at each instant of time or at a particular point in distance is directly proportional to the then existent magnitude or amount of whatever there is that is attenuating.

In the code of differential calculus this fact may be written $y = e^{-ct}$ where c is the decrement constant

$$\frac{dy}{dt} = -ce^{-ct} = -cy. \quad \text{See } \frac{dy}{dx} \text{ curves of chart, page 103, Fig. 43.}$$

This means that the slope of the attenuation or decay curve $Y = e^{-ct}$ varies as one goes along the curve and is $-cy$. The slope is greater at the beginning when Y , the decaying factor (whatever it may be) is large. The

slope gradually decreases. The Y might represent a number of milliwatts of power at some point along a long distance telephone line or the temperature of a hot object which is cooling off as a function of time.

One might think that there seems to be a contradiction in the two basic truths about geometric retrogression or attenuation.

The ratio between successive values in such a decay is a constant if one considers values at two discrete different periods of elapsed time or at two discrete different distances along a telephone wire line. There is no contradiction however. It is true that the instantaneous rate of change is directly proportional to the magnitude of the factor which is attenuating, at a particular instant of time or at a particular point along the circuit. This means again that the instantaneous rate of change decreases as the absolute magnitude decreases.

These two facts are closely knit associated principles. One might think of the latter being the truth because the former is true.

When in this treatise more complete consideration is given to the attenuation of power along a telephone line, the two characteristics of geometric retrogression will be called upon to acquire good understanding.

SOME MORE SIGNIFICANT EXAMPLES OF GEOMETRIC RETROGRESSION IN NATURAL PHENOMENA

Reference has been made previously in this treatise to the decay of electrical quantities when capacitors and inductances are de-energized. In several charts shown later, see pages 186, 191, 199 electrical transients are portrayed. In all the situations portrayed by curves in these charts, the decay occurs in geometric retrogression fashion. The reader might take a look at the charts now even before he wishes to study them carefully.

In Fig. 74 the equation is $q = Qe^{-\frac{t}{RC}}$

If one uses the time constant (T), namely RC , as a unit of time, then the equation becomes $q = Qe^{-\tau}$.

After T seconds, the q is $.368Q$. After $2T$ seconds, the q is $.135Q$ or $(.368 \text{ times } .368)Q$. The .050, .018 and .007 values shown in Charts are $(.368)^2$, $(.368)^3$ and $(.368)^4$. If T were 10, $q = .000045Q = .368^{10}Q$.

One could write the above equation in the basic form $q = Q\left(\frac{1}{e}\right)^{\tau} = Q \cdot .368^{\tau}$.

The constant ratio between consecutive terms for equal intervals of time (T) is $\frac{1}{e}$ or $.368 = e^{-1}$

One could, if it were one's choice, write the basic equation in the following form

$$q = Q\left(\frac{1}{10.43}\right)^{\tau} = Q10^{-.43\tau} = Qe^{-\tau}$$

The same ideas and equations apply to the voltage and current decay curves in the other transient curve Charts, pages 181-200.

GEOMETRIC PROGRESSION IN ELECTRICAL CIRCUITS

As has been pointed out in preceding paragraphs, the decaying curves of voltage, current and charge portray geometric retrogression in every case. It is interesting to realize that, contrary to one's expectation, the growth or rise of all these quantities does not take place in geometric progression fashion. The growth equations involve 1 minus a factor, as, for example:

$$q = Q(1 - e^{-\frac{t}{RC}}) = Q(1 - e^{-\tau}) = Q(1 - .368^{\tau})$$

The second term in the parenthesis is a natural decay factor but one (1) minus this natural decay factor does not produce a natural growth process. This says that the ratio between 1, .368, .135, .050, .018 and .007 is a constant, namely $.368$ or $\frac{1}{e}$. The ratio between 1 minus these quantities, namely between .632, .865, .950, .982 and .993, is not a constant.

If one holds one of the transient charts (Fig. 70, page 194) up to light upside down and looks through the Chart from the back, the natural growth curve becomes a natural decay curve.

NEWTON'S LAW OF COOLING

A hot body cools off by radiation. The temperature decreases with elapsed time as per the natural decay doctrine.

The basic equation is $T = T_0 \cdot .95^t$ where T_0 is the original temperature of body above the surrounding medium and t is time in minutes and .95 is the cooling decay constant ratio characteristic of a particular hot body in a particular environment. In a particular case, if T_0 were 72° , then the hot body would cool to 3.31° in 60 minutes, which means .95 to the 60th power is .046

$$72^\circ \text{ times } .046 = 3.31^\circ$$

The three exponential equations would be

$$T = T_0 \cdot .95^t = T_0 \cdot e^{-.0012t} = T_0 \cdot 10^{-.0022t}$$

$$\text{If } t = 60 \text{ minutes and } T_0 = 72^\circ$$

$$T = 72^\circ \cdot .95^{60} = 72^\circ \cdot e^{-.072} = 72^\circ \cdot 10^{-.332} = 72^\circ \cdot .046 = 3.31^\circ$$

Again, the above 3.31° is easily calculated by use of logarithmic tables for either of the two logarithmic bases named in the equations.

DECREASE OF ATMOSPHERIC PRESSURE AT HIGH ALTITUDES

It is a simple and well known fact that as one goes up in the air (literally, not figuratively), the atmospheric pressure decreases. It decreases as per the natural decay doctrine. The equation under average conditions $P = P_0 \cdot .86^h$ where " h " is in kilometers. The .86 is the constant ratio of decrease of pressure per kilometer. The three equations are

$$P = P_0 \cdot .86^h = P_0 \cdot 10^{-.0645h} = P_0 e^{-.1508h}$$

At a height of 50 kilometers, the pressure (barometric height) would be .32 millimeters assuming $P_0 = 760$ millimeters at the earth's surface or sea level. At 20 kilometers, the pressure would be 42 millimeters and at 10 kilometers the pressure would be 199 millimeters. Actually of course, the constant of decrement in the three equations is not always the same under different weather conditions. However in general, one can validly say, that the logarithm of the air pressure varies inversely proportional to the height. The logarithmic decrement constants named are correct to a small percentage error.

If h is in miles then $P = P_0 \cdot .78^h = P_0 e^{-.2327h} = P_0 \cdot 10^{-.1097h}$.

One mile = 1.609 kilometers

LIGHT ABSORPTION

When light passes through a piece of clear or colored glass, the light intensity transmitted decreases with thickness as per the natural decay doctrine.

For a particular kind of glass, the three equations

$$I = I_0 \cdot .9802^t = I_0 \cdot 10^{-.0087t} = I_0 \cdot e^{-.0204t}$$

where t is the thickness in centimeters. For a piece of glass 10 cms thick, the I would be 82 % of I_0 and for 35 cms I would be 50 % I_0 .

In all cases of the many natural phenomena, the constant in the exponent of the latter two forms of the equation can be appropriately called the decrement constant or the logarithmic decrement.

TUNING FORKS AND VIBRATING STRINGS

The amplitude of vibration of a tuning fork and harp string, for example, decreases with increasing elapsed time as per the natural decay or geometric retrogression doctrine. One can feel confident that the amplitude decay does occur in geometric retrogression fashion since they are fine examples of natural decay—a natural

phenomenon on its own. This particular case is explored and examined in detail later. It is suggested to the reader that right now he take a look at two Charts, pages 116 and 117, for a beautiful portrayal of decreasing displacement for not only peak values (amplitude) but also for intermediate instants of time. Figs. 46, 47

NEGATIVE EXPONENTS

It has been pointed out and emphasized that at least three exponential equations can be written for geometric retrogression phenomena. They are essentially the same equation. If the logarithmic bases 10 and "e" are used, then a negative sign conventionally appears in the exponent. Basically the exponent is not negative but is positive. The important fact is that the ratio between any two successive values of the factor involved is less than unity. The varying factor, which may be time or distance, etc., is of course the positive exponent.

The equation for discharge of a capacitor is written

$$q = Qe^{-\frac{t}{RC}} = Q10^{-.434\frac{t}{RC}}$$

It is perhaps better to write it

$$q = Q\left(\frac{1}{e}\right)^{\frac{t}{RC}} = Q.368^{\frac{t}{RC}} = Q\left(\frac{1}{10^{.434}}\right)^{\frac{t}{RC}}$$

Again it is to be said that the exponent basically always has a positive sign whether the ratio is greater or less than unity to specify a geometric progression or retrogression, respectively.

DECREASE OF AMPLITUDE OF AN ELECTRIC CURRENT ALONG A CIRCUIT

Ordinarily a sine wave of alternating electric current in a circuit of a few feet or yards in length remains constant in amplitude or value in amperes throughout the circuit.

However, in a telephone circuit of several or hundreds of miles, the amplitude of the current decreases as per the doctrine herein discussed. The equation ordinarily would be $i = I_0 \sin \omega t$ where I_0 is the maximum value of current or amplitude and t is the elapsed time.

If a sine wave of alternating current were traveling out along a circuit with a velocity of V , then the t involved for the sinusoidal current at some distance X would be $\frac{X}{V}$.

We can then write

$$i = I_0 \sin \omega t = I_0 \sin \omega \frac{X}{V}$$

However, as stated above, the amplitude I_0 decays as the current continues along the circuit. Then $I = I_0 r^X$ where " r " is a ratio less than unity, and is the ratio of values of amplitude at points one mile apart. X is the number of miles.

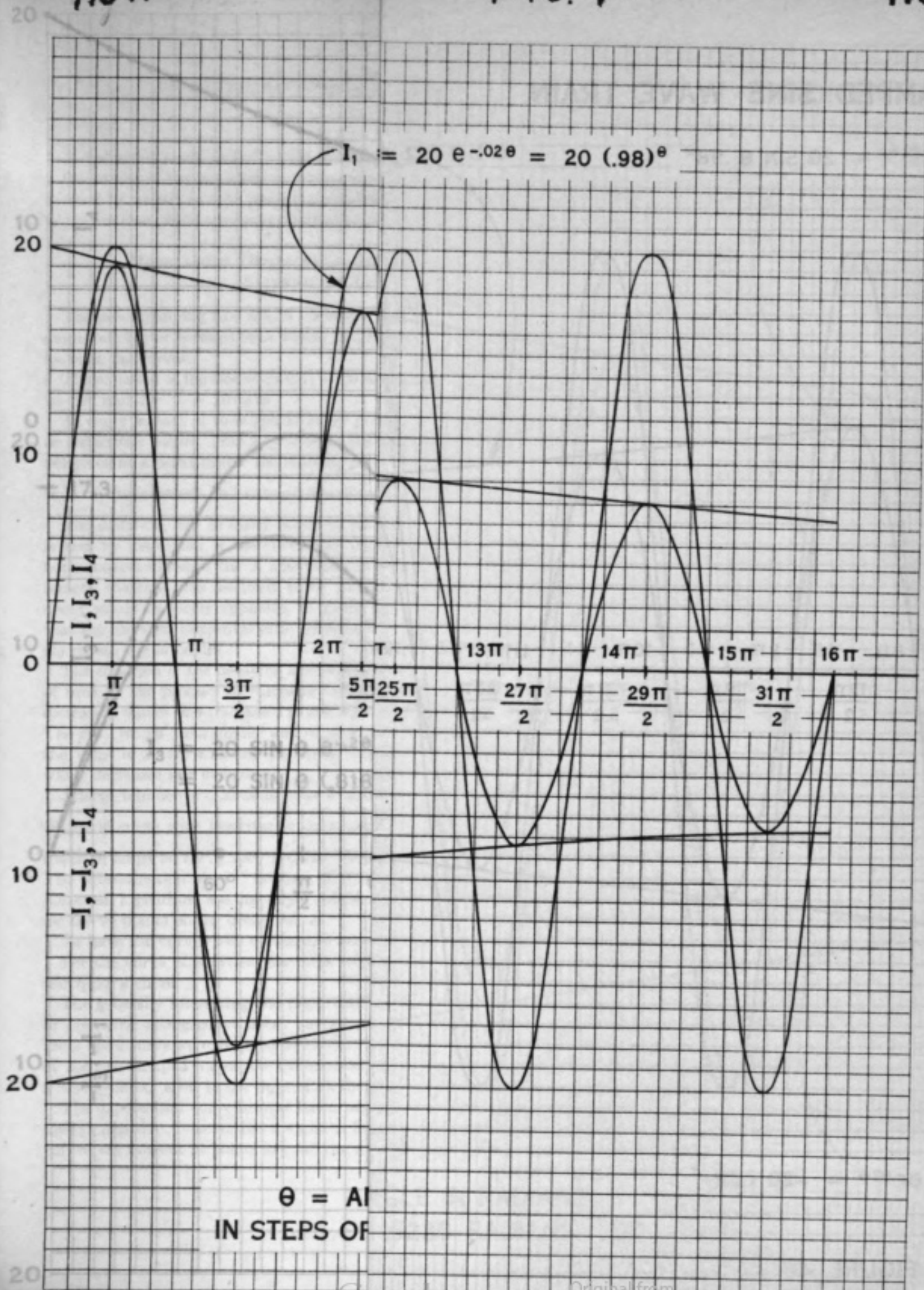
Assume the value of " r ", the constant of decrement is .9 for a particular type of circuit. Then $I = I_0(.9)^X$ would be the basic equation for the amplitude decay per se. The exponent X which is an angle to be expressed in radians would be in steps of 2π radians after the first step of $\frac{\pi}{2}$ (90°).

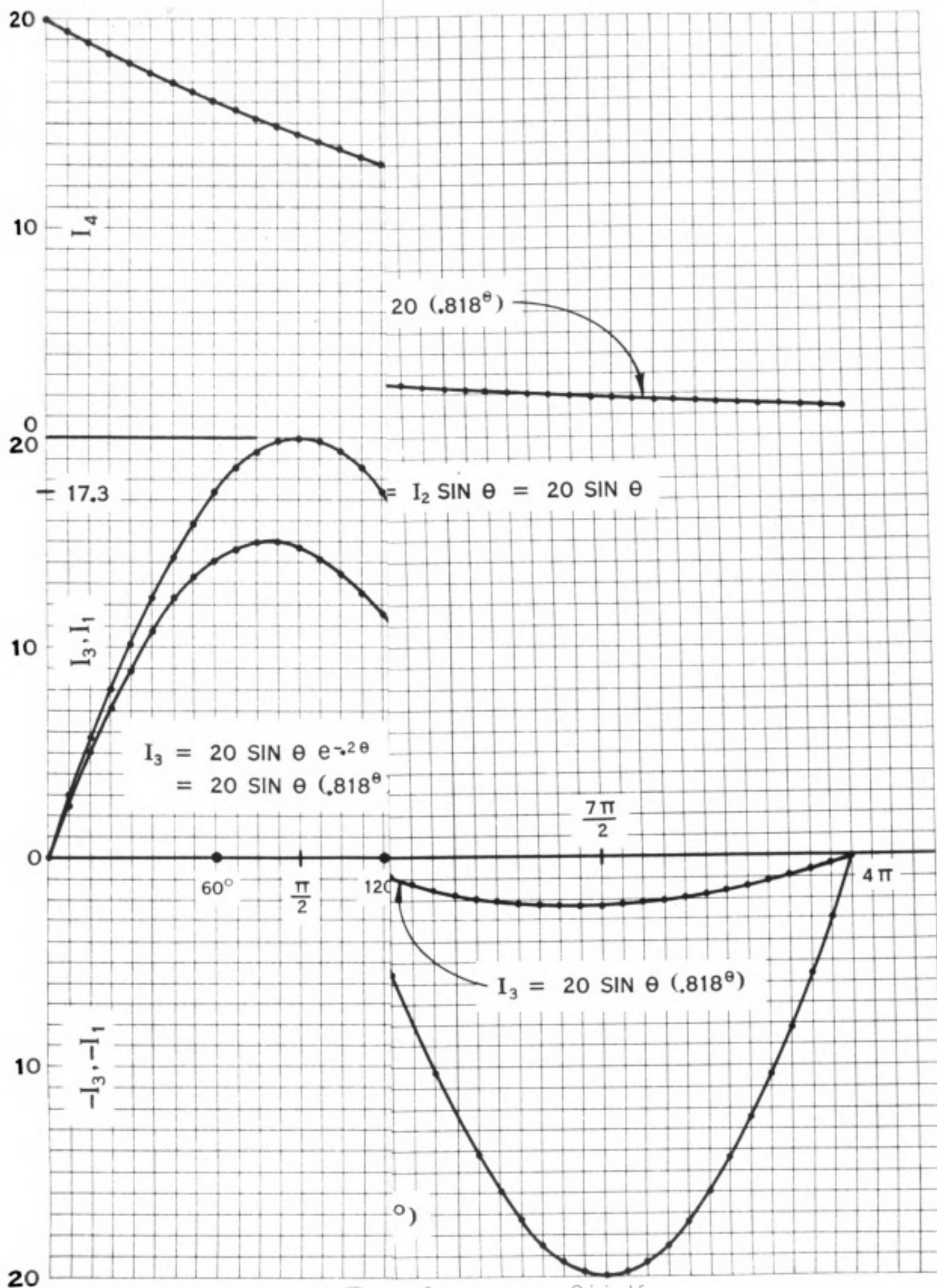
Then in turn, $i = I_0(.9)^X \sin \omega \frac{X}{V}$ would be the complete equation.

If we choose to write the equations in terms of the two conventional logarithmic bases then

$$\begin{aligned} i &= I_0 10^{-.0434X} \sin \omega \frac{X}{V} \\ &= I_0 e^{-.1054X} \sin \omega \frac{X}{V} \end{aligned}$$

The following accompanying curves graphically portray the diminishing amplitude I_0 and varying values of current at individual values of X .





SUMMARY

A statement made earlier merits restatement.

Quantities of various kinds occurring in natural phenomena grow and decay. It is significant that these are decreed by nature to take place in geometric progression and geometric retrogression fashion. It is interesting to realize that there are many more examples of decay or attenuation than growth.

MORE ABOUT DECREASE OF AMPLITUDE OF TUNING FORKS, VIBRATING STRINGS AND
ALTERNATING CURRENT ALONG AN ELECTRICAL CIRCUIT

The accompanying two sets of curves (Figures 46 and 47) and their equations tell the story of attenuating amplitude better perhaps than words. Before reading and studying further, it is suggested to take several good looks at the curves.

The equation of the sine wave is $I_1 = 20 \sin \theta$. The equations of the two exponential curves are $I_4 = 20e^{-.02\theta} = 20 \cdot .98^\theta$ and $20e^{-.3\theta} = 20 \cdot .818^\theta$.

At every instant of time and at every position out on the electric circuit, whatever the case may be, a levy or reductions occurs. This levy reduces the tuning fork displacement or the current which existed an instant before or at a point nearer the starting point of the circuit. It has been assumed that the unimpaired amplitude of displacement and current was 20 units at $\frac{\pi}{2}$ (90°) radians.

The situation, like so many others in nature, is one where the law of geometric retrogression prevails. So it is seen by inspection of the curve showing eight (8) complete cycles that the amplitudes (maximum value) both above and below the X axis touch an exponential curve ($I_4 = 20e^{-.02\theta} = 20 \cdot .98^\theta$). The logarithmic decrement constant was assumed to be .02. This provided a rather substantial rate of attenuation so that the decay would be nicely apparent. It meant that the attenuation for 1 radian would be .98. This would mean that the constant numerical ratio between any two successive values of current when the angle differed by one (1) radian would be .98.

The equation of the decaying sine wave curve ($I_3 = 20 \sin \theta e^{-.02\theta}$) means that I_3 is obtained at any instant of time by the product of values of two curves. The value of $\sin \theta$ itself alternates in its rises and falls with increasing values of θ (radians). It is a repetitive and cyclic curve. The exponential curve is not a cyclic curve. The $e^{-.02\theta}$ or $.98^\theta$ gets smaller and smaller with increasing value of the angle θ . To get the decaying sine wave, the value of $20 \sin \theta$ at any particular value of θ is to be multiplied (thereby reduced) by an amount $e^{-.02\theta}$ which decreases as θ increases.

A large number of calculations were made to get the two sets of curves. For every point or dot on the two (2) cycle curve with logarithmic decrement constant of .2, ($e^{-.2} = \frac{1}{e^{.2}} = \frac{1}{1.2214} = .818$) a calculation was made in steps of 7.5° (.1309 radians). The plotting of $I_1 = 20 \sin \theta$ in steps of 7.5° was not too arduous. The determination of $e^{-.3\theta}$ in steps of .1309 radians for θ for all the points was a simple task. A fine table of Napierian logarithms was happily available. The slide rule was used to do the multiplying of $20 \sin \theta$ and $e^{-.3\theta}$ for the 96 points on the decay curve.

To have the curves well outlined, it was necessary to calculate the many individual points on them.

For the curve of 8 cycles, the steps were 30° (.5236 radians). The total number of calculations (again 96) was quite a chore.

The generally interesting and important aspects of the two sets of curves is that the two exponential curves or geometric retrogressive decay curves are the envelopes of the damped curves. If the exponential curve of the 2 cycle curve were dropped down the right amount, it would be the envelope of the damped curve. In each of the two damped waves, the exponential curve would touch and be tangent to the damped curve at points a bit beyond or a bit to the right of all the positive maxima.

The exponential curve always has a negative slope. Its negative slope is not constant to be sure but it is always negative. It never has a segment however small which is horizontal. It is imperative therefore that the curves are tangent at points beyond the maxima because at the maxima of the damped curve, there is a tiny

segment which is actually horizontal. At these horizontal tiny segments, the slope is zero. One must go a bit beyond the positive maxima to encounter negative slopes.

SOME ILLUSTRATIVE SPECIFIC CALCULATIONS

In order to help the reader better understand and to appreciate the fact that the decaying amplitude curve is the result or produce of values in two curves (a sine curve and an exponential curve), attention will be called to two points on the two cycle curves. The points involved are those directly above the 60° and 120° points on the X axis.

As has been the practice in this book, the geometric retrogression idea has been expressed by three equations. The first equation embodies and nicely states the simple and fundamental concept, $I_4 = 20 \cdot .818^4$. The second two equations give directions as to how to get an answer conveniently, use logarithms either to the base e or the base 10. So above equation becomes $I_4 = 20e^{-.36} = 20 \cdot 10^{-.060908}$.

For sake of illustrating the first equation, we will do without logarithms. Consider the value of $\sin 60^\circ$; which is eight (8) steps of .1309 radians each to the right of the origin 0. The $\sin$ of 60° is .866 which when multiplied to 20 provides a point on the undamped curve of 17.3 as denoted by arrow. Now 60° is close to one radian (57.3°). So if the exponent of the constant ratio of .818 is 1, then $.818^1$ is .818. Now 17.3 times .818 = 14.3 as the value of I_3 . Note arrow on this curve. The point on curve is 14.0; a bit less. However 60 degrees is 1.0472 radians and $.818^{1.0472} = e^{-.2 \cdot 1.0472}$ or .810 instead of the .818. So $20 \cdot \cos 60$ (17.3) multiplied by $.818^{1.0472}$ or by $e^{-.20942}$ or by .810 gives 14.0. However without use of "logs," one gets a pretty close value, namely 14.3.

For 120 degrees which is close to two radians ($114^\circ 36'$), the sine is also .866 which when multiplied by 20 provides a point on undamped curve of 17.3 as denoted by arrow. Now $.818^2$, for the point on the curve when θ is 2 radians, is .669(.7). So now the 17.3 must be multiplied by a lesser value than before (.7 compared to .818), which gives 12.1. The point on curve has an ordinate of 11.4. The values are near enough to illustrate the idea of the simple equation, $I_2 = 20 \sin \theta \cdot .818^2$.

Since 120° equals 2.094 radians, one does well to call on logarithms. $.818^{2.094} = e^{-.2 \cdot 2.094} = e^{-.4188} = .66$. So to get a precise answer for the value of I_2 at 120° , one multiplies $20 \sin 120^\circ$ (17.3) by .66. The answer is 11.4 (as compared to 12.1) which is value used to get curve.

For the many points on curve involving the steps of 7.5° (.1309 radians), one wisely invokes logarithms.

Each point on the damped curve represents an interval of time after a tuning fork or a vibrating string was activated with some initial amplitude. At later instants of time, the individual displacements throughout the swing decay. If one considers the maximum displacements or amplitudes at $T/4$ and $5/4T$, (where T is period), they decrease and their decaying values touch the two envelope exponential curves.

Each point on the curves may represent the situation of decaying current out at some distance on a long electric circuit to which the current has traveled in a certain instant of time.

The decaying current at that point and every point in the long circuit at a particular instant of time is portrayed by accompanying curves. Even though the logarithmic decrement constants are seemingly numerically small, the decay in one case is very rapid and in the other case (.02) the decay is 50 percent in 5.5 cycles.

NUMBER OF OSCILLATIONS IN A DAMPED WAVE TRAIN

There are theoretically a great many oscillations in every discharge of a capacitor connected in a circuit with resistance and inductance. In a practical sense, the wave train can be considered ended when the oscillations are reduced to a negligible magnitude. The fraction of the last amplitude compared to the first amplitude that can be considered negligible depends on the use to which the oscillatory circuit is put. When the last amplitude is reduced to 0.01 of the initial amplitude, a reasonable degree of negligibility is assumed. However in these days of high power radio transmitter outputs and sensitive receivers, one might better assume a ratio of .001, .0001 or .00001. After all in a radio signal of 10^{-10} milliwatts the amplitude of alternating voltage in that signal is very, very small. (Unless masked by noise, the very tiny signals are most effective.)

It would be interesting to find out how many oscillations there would be in the case of the two damped wave trains plotted in accompanying charts, Figures 46, 47. The number obtained depends of course on the criterion of negligibility assumed.

In Figure 46, the constant ratio between successive positive and successive negative amplitudes is 1.12. It is to be pointed out again that the amplitudes are decreasing in geometric retrogression fashion.

As per discussion in several previous pages of this treatise, one can understandingly write down $100 = r^n$, where r is the constant ratio between successive amplitudes, the 100 is the assumed ratio of initial amplitude to last amplitude and n is the number of oscillations. As before, the method of attack is to write this basically simple equation in terms of logarithms

$$\log_e 100 = n \log_e r$$

$$n = \frac{\log 100}{\log 1.12} = \frac{4.6}{.12} = 38 \text{ oscillations}$$

In turn as per first basic equation, this means

$$100 = 1.12^{38}$$

If the criterion of negligibility were 1000

$$n = \frac{\log 1000}{\log 1.12} = \frac{6.9}{.12} = 57.5 \text{ oscillations}$$

If the criterion of negligibility were 10,000

$$n = \frac{\log 10,000}{\log 1.12} = \frac{9.25}{.12} = 77.0 \text{ oscillations}$$

$$10,000 = 1.12^{77}$$

The data for Figure 47 would be based on a constant ratio of 3.5 for successive amplitudes.

$$n = \frac{\log 100}{\log 3.5} = \frac{4.06}{1.26} = 3.6 \text{ oscillations}$$

$$n = \frac{\log 1000}{\log 3.5} = \frac{6.09}{1.26} = 5.4 \text{ oscillations}$$

$$n = \frac{\log 10,000}{\log 3.5} = \frac{9.25}{1.26} = 7.3 \text{ oscillations}$$

The ratio of any positive amplitude to the next negative amplitude and the ratio of any negative amplitude to the next positive amplitude would be the $\sqrt{1.12}$ or 1.06 (Fig. 46). In Figure 47, the corresponding value would be $\sqrt{3.5}$ or 1.87.

The ratios of amplitudes in Fig. 46 and 47 are 1.12 and 3.5 respectively. The logarithms to the Napierian base (e) are .12 and 1.26 respectively. The logarithm of the ratio in any wave train is called logarithmic decrement in contrast to logarithmic decrement constant as this latter term has been used in previous pages. Logarithmic decrement constants have been designated by C and K . The logarithmic decrement is designated by the Greek letter delta, δ . The ratio relates to two successive positive or two successive negative amplitudes.

EXPONENTIAL LOUD SPEAKER HORN

If one looks at the 8 cycle envelope exponential curves from right to left, one sees the contour of an exponential loud speaker horn. Reference to this idea was made in previous pages when some ideas and calculations for exponential horns were presented. See page 108, Figure 45.

AN EXPONENTIAL PROBLEM IN MECHANICS

The Power Driven Winch or Capstan

An interesting and splendid example of exponential change or growth is found in the winch or capstan used for hoisting heavy loads aboard ship, for instance. As is to be seen from the accompanying sketch, Figure 48,

the winch mechanism consists of a power driven drum around which a rope is wound. A workman applies a force at the slack portion of the rope. Since the rope is wound around the drum, the magnitude of the applied force, F_A , is greatly increased by the friction of rope and drum surface to produce a greater force F_L . The energy required to lift the "load" is usually supplied by an electric motor turning the drum. In some cases the drum is operated by a crank system turned by other workmen. Of course, some work is done by the first workman. The important fact, however, is that the force F_A is greatly increased to an amount required to lift the heavy load, F_L .

It is the relation of the magnitude of the two forces, F_A and F_L , that will be herein discussed. The rope makes contact with the metal or wooden surface in varying amounts. The amount of contact is to be considered in terms of angle, measured in radians, subtended by the rope at the axis of rotation of the circular drum. As the amount of angle is increased, the magnitude of F_L increases with an assumed fixed value of F_A .

In the second accompanying sketch (Figure 49) the situation is portrayed for three conditions. The angle of contact (θ) is one radian (57.3°) in one case. In the second case the angle θ is $\frac{\pi}{2}$ radians. In the third case the angle θ is 2π radians, since the rope makes one turn around the drum. For two turns the angle would be 4π radians; for three turns the angle would be 6π radians.

It is an interesting and amazing fact that the numerical relation between the two forces is given by the law of geometric progression which, in turn, means an exponential change.

$$F_L = F_A K^\theta$$

In this equation K is a constant depending on the coefficient of friction and θ is the angle of surface contact of rope and drum. The equation can be changed so as to involve the coefficient of friction C and is as follows:

$$F_L = F_A \cdot e^{C\theta}$$

Here again is a natural phenomenon or natural situation governed by the basic truth of geometric progression or exponential relation.

It is now proposed to illustrate this problem by use of numerical values which will bring out the surprising amounts by which F_L exceeds F_A .

COEFFICIENTS OF FRICTION = C

The coefficients of friction between various surfaces must of course be less than unity. In general, experiment reveals the coefficients of friction between various rope materials, various woods and various metals vary between .20 and .55. These values for C are assumed in the following calculations.

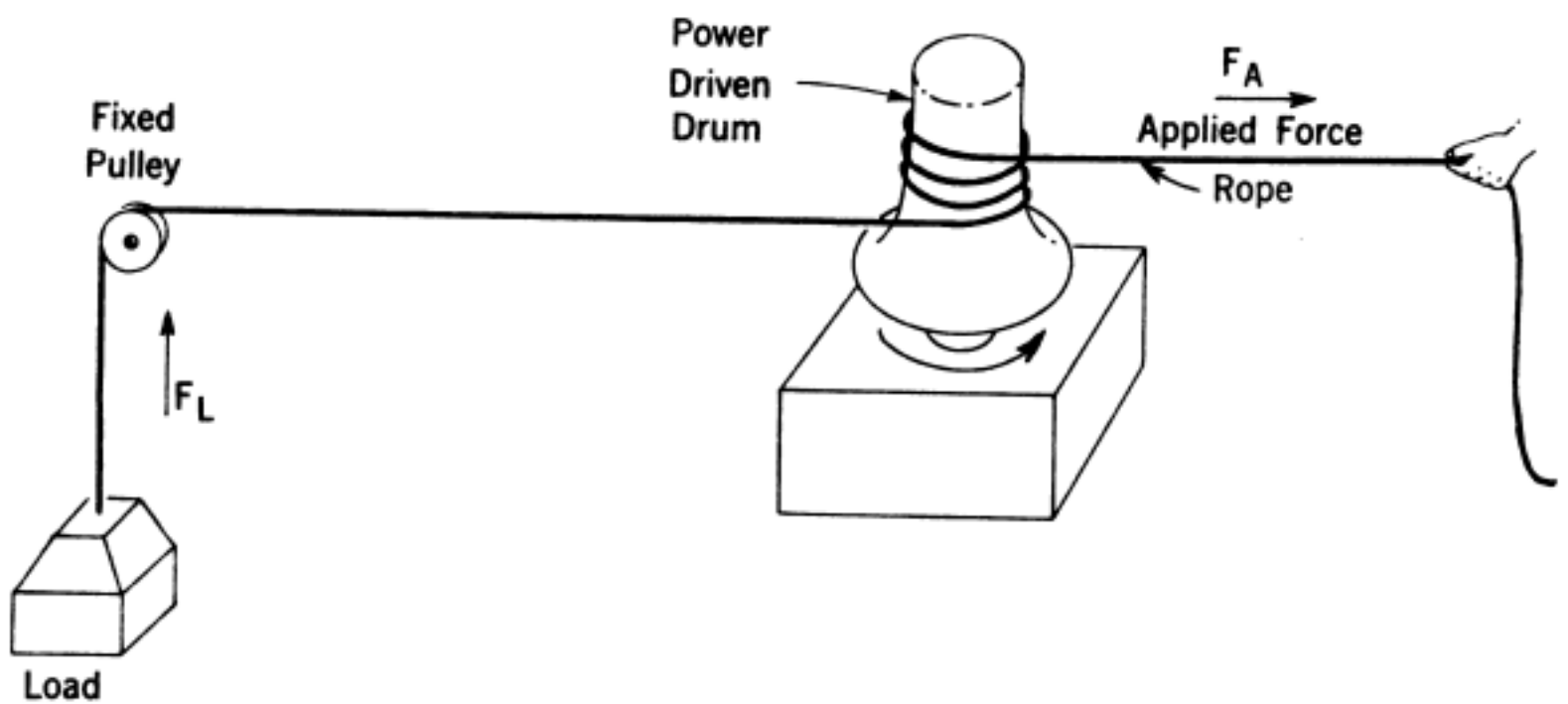
It is in order to find the numerical value of e^C for the assumed varying values of C to determine the various values of K ; θ will be unity (1) radian.

$e^{.20} = 1.22$	$e^{.40} = 1.49$
$e^{.25} = 1.28$	$e^{.45} = 1.57$
$e^{.30} = 1.35$	$e^{.50} = 1.65$
$e^{.35} = 1.42$	$e^{.55} = 1.73$

In one case, then, the basic equation illustrating the simple truth could be $F_L = F_A 1.65^\theta$. This would mean that if the angle of contact (θ) were 1 radian, and C were .50, F_L would be 1.65 greater than F_A . If θ were 2 radians, F_L would be 1.65² or 2.72 greater than F_A . If θ were 3 radians, F_L would be 1.65³ or 4.49 greater than F_A . If θ were 2π radians (one turn), F_L would be 4.81 greater than F_A .

In the practical use of the winch, one would have, of course, one or more turns of the rope around the drum, involving 2π , 4π , 6π , etc. radians corresponding to 1, 2 and 3 turns, respectively.

It is in order, therefore, to calculate the numerical value of $e^{C \cdot 2\pi}$, $e^{C \cdot 4\pi}$ and $e^{C \cdot 6\pi}$ for the assumed values of C .



WINCH SYSTEM
FIGURE 48

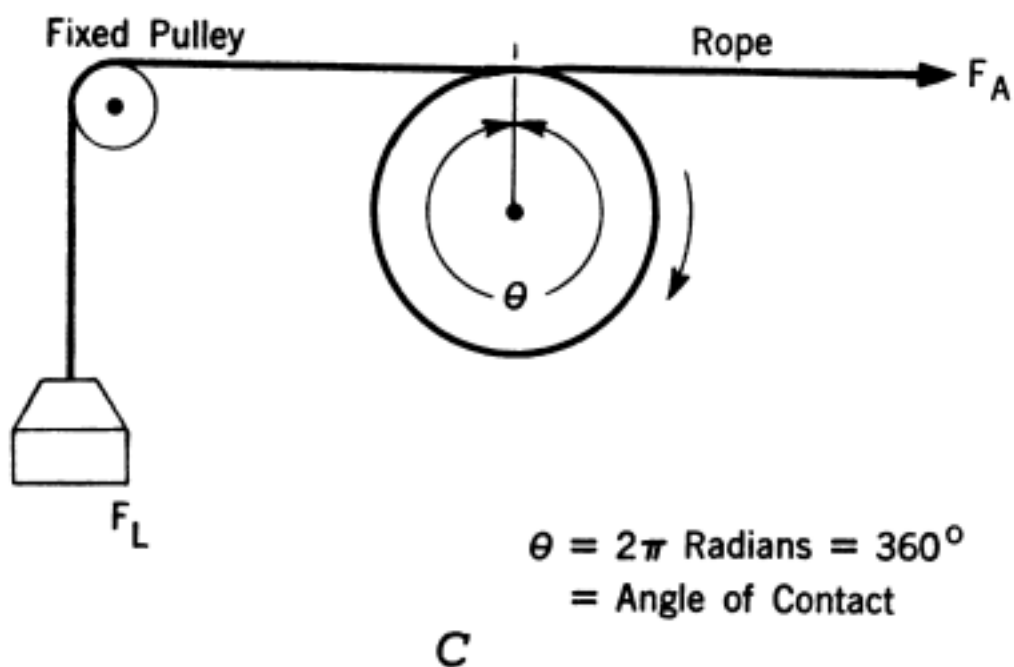
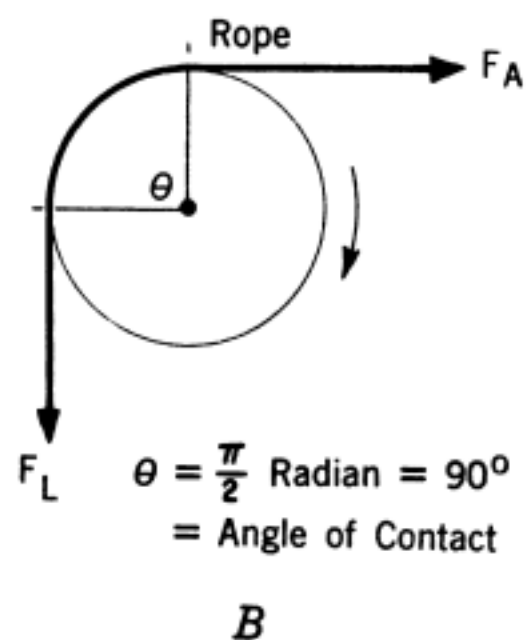
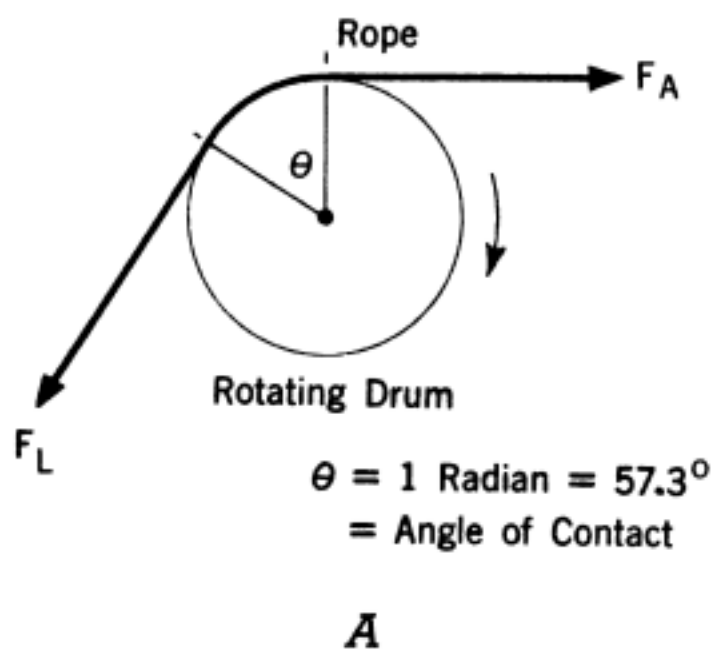


FIGURE 49

One Complete Turn of Rope

$$\frac{F_L}{F_A}$$

$$\begin{aligned} e^{.2 \cdot 2\pi} &= e^{1.26} = 3.53 \\ e^{.25 \cdot 2\pi} &= e^{1.57} = 4.81 \\ e^{.30 \cdot 2\pi} &= e^{1.88} = 6.55 \\ e^{.35 \cdot 2\pi} &= e^{2.20} = 9.03 \\ e^{.40 \cdot 2\pi} &= e^{2.51} = 12.31 \\ e^{.45 \cdot 2\pi} &= e^{2.83} = 16.95 \\ e^{.50 \cdot 2\pi} &= e^{3.14} = 23.10 \\ e^{.55 \cdot 2\pi} &= e^{3.46} = 31.82 \end{aligned}$$

Two Complete Turns of Rope

$$\frac{F_L}{F_A}$$

$$\begin{aligned} e^{2.4\pi} &= e^{2.52} = 12.43 = 3.53^2 \\ e^{2.5\pi} &= e^{3.14} = 23.10 = 4.81^2 \\ e^{3.0\pi} &= e^{3.76} = 42.95 = 6.55^2 \\ e^{3.5\pi} &= e^{4.40} = 81.45 = 9.03^2 \\ e^{4.0\pi} &= e^{5.02} = 151.4 = 12.31^2 \\ e^{4.5\pi} &= e^{5.65} = 287.5 = 16.95^2 \\ e^{5.0\pi} &= e^{6.28} = 533.6 = 23.10^2 \\ e^{5.5\pi} &= e^{6.92} = 1012.5 = 31.82^2 \end{aligned}$$

Three Complete Turns of Rope

$$\frac{F_L}{F_A}$$

$$\begin{aligned} e^{2.4\pi} &= e^{5.0} = 43.88 = 3.53^3 \\ e^{2.5\pi} &= e^{6.28} = 111.11 = 4.81^3 \\ e^{3.0\pi} &= e^{7.62} = 281.32 = 6.55^3 \\ e^{3.5\pi} &= e^{8.90} = 735.49 = 9.03^3 \\ e^{4.0\pi} &= e^{10.04} = 1,863.7 = 12.31^3 \\ e^{4.5\pi} &= e^{11.32} = 4,873.1 = 16.95^3 \\ e^{5.0\pi} &= e^{12.56} = 6,990.2 = 23.10^3 \\ e^{5.5\pi} &= e^{13.84} = 32,217.8 = 31.82^3 \end{aligned}$$

In accordance with the last calculation, a pull of one pound on the part of a workman would produce a pull of 32,218 pounds. To lift a load of this amount 10 feet, the electric motor would supply 322,170 foot-pounds of work while the workman would do only 10 foot-pounds of work. One is reminded of a vacuum tube amplifier and power supply in electrical systems. A small electrical input is amplified to a large electrical output. The difference in forces can be compared to the different voltages of the input and output circuit of the amplifier. The power supply of the amplifier is the source of renewed power.

In the actual operation of the power driven winch, the workman continuously pulls on his "free" end of the rope and permits it to fall on the platform or ship deck. When the load is lifted by the electric motor to the desired height, the workman releases his pull, whereupon the drum rotates within the loose turns of the rope around it.

It is to be observed that the increase of Force, F_L , achieved by wrapping the rope twice around is very substantial as compared to one turn. The three-turn wrapping gives an equal ratio of increase over two turns. The increase is exponential.

Even if the coefficient of friction were small, namely from .20 to .35, the addition of a fourth complete turn of the rope would provide an $\frac{F_L}{F_A}$ of 154.90 (3.53^4), 534.44 (4.81^4), 1,842.65 (6.55^4) and 6,740.5 (9.03^4) respectively.

If one turn is taken as a basis for the equation, the exponent in the equation becomes equal to the number of turns. In one illustrative case, the equation can be written for a coefficient (c) of .4 as follows:

$$F_L = F_A \cdot 12.31^N$$

where N = number of complete turns; also when $c = .55$ and $N = 3$,

$$F_L = F_A \cdot 31.82^3 = 32,217.8 F_A$$

This numerical value also illustrates that a workman can wind a rope three or four times around a circular post or cleat on a big barge and thereby keep the barge securely anchored to a fixed post on a pier. By exerting

a small force on his end of the rope, he readily keeps the barge "to shore". A terrific gain in force is obtained by the winch system. The exponential law fortunately applies and controls the situation.

NATURE'S DECAY LAW OF GEOMETRIC RETROGRESSION APPLIED TO POWER IN ELECTRICAL TRANSMISSION OF VOICE POWER

We are moving along on the way to getting an answer to the question propounded early in this treatise. We have the exponential expression e^x and e^{-x} . The concluding steps in getting an answer to the question at hand are to coordinate and maneuver the various tools, ideas and equations at hand so that they will nicely provide a rational and understandable conclusion.

It is in order therefore to consider the practical case of a telephone circuit along which electrical power is carrying speech. Out of an analysis of the quantitative aspects of this problem, we will get some more ideas and principles by which we can achieve the objective.

In the first place, we will accept the simple fact that nature's law of decay, namely the law of geometric retrogression, applies to this case as it does to many other types of natural phenomena.

One does not need to be reluctant to accept, on faith as it were, the doctrine of geometric retrogression as a basic principle. There need be no misgivings as to its validity and applicability. It is so simple and seems to be too good and too easy to be true. It is not abstrusely mathematical. It provides a fundamental approach with which to study the quantitative aspects of many and various processes in their decays. See curves of Figure 50.

We will assume that in the particular cable circuit under discussion, the constant decay ratio between the power at every pair of two points one mile apart is .871. Actual precise measurements with proper testing meters suggest this particular ratio of .871 as a characteristic of certain cables.

At once then we can write the equation for the attenuation of power along the circuit as the power decreases with increasing distance. The equation is $P = P_0 (.871)^d$

This equation says that the power (P) at a distance of d miles is some fraction of the original power P_0 . The ratio for any and every two points one mile apart is .871. At a point 5 miles from the starting point out on the circuit, the power is $P = P_0 (.871)^5$.

Now .871 raised to the fifth power is $\frac{1}{2}$ or .5. This means that .871 multiplied by itself 5 times is $\frac{1}{2}$ or .5. The reader might wish just for fun to check it by sheer multiplication.

At the end of the second five miles, the power also will be .5 of the power at the end of the first 5 miles.

If one assumed that P_0 was 1000 milliwatts, then in 5 miles the power is 500 milliwatts. At the end of 10 miles it will be .25 of 1000 or 250. This means that .871 raised to the 10th power is .25. So the power along this circuit attenuates at the .871 rate per mile. At the end of 50 miles the power will be 1 milliwatt since $(.871)^{50}$ is .001.

To be sure, there will be slight numerical variations of the $\frac{1}{2}$ of $\frac{1}{2}$ for every 5 miles but in a general way, the answers come out very close numerically as seen in an accompanying Table, Figure 51.

THREE WAYS TO WRITE DECAY EQUATION

The simple and readily understandable way to write the natural attenuation curve is as above $P = P_0 (.871)^d$

To use this equation as is to calculate P at some distance, say 40 miles, would involve a lot of sheer and tedious multiplication. So as discussed earlier let us use logarithms to the base 10 and to the base e . The base of the logarithms in the above exponential equation is .871 and the logarithm is d , the number of miles. Since there are no log tables available with .871 as a base, we can readily change to the base 10, or to the base e . As before illustrated we ask, what is logarithm of .871 to base 10? It is easily found in log tables as $-.06$.

It might be helpful to observe that

$$\begin{aligned} .871 &= \frac{1}{1.14816} \\ &= \frac{1}{10^{.06}} \\ &= 10^{-.06} \end{aligned}$$

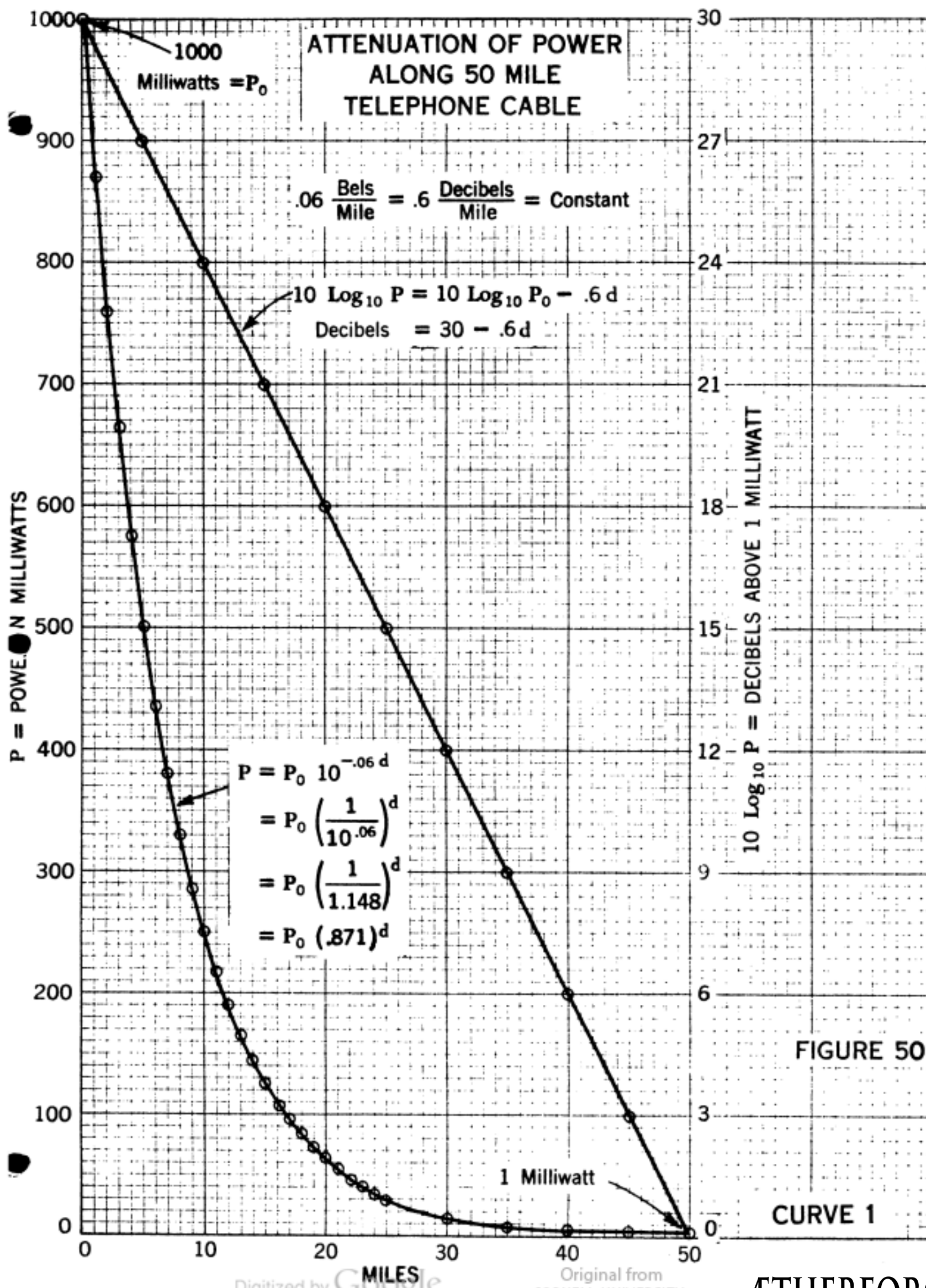


TABLE 1

Some Calculations for a Telephone Cable Circuit

Using Logarithms to the Base 10

$$p \approx p_0 (.871)^d = p_0 \left(\frac{1}{1.14815} \right)^d = p_0 \left(\frac{1}{10.06} \right)^d = p_0 \frac{1}{10.06^d} = p_0 \cdot 10^{-.06d}$$

.00 Bel per Mile = .6 Decibels per Mile Power Attenuation Characteristic
of Circuit $\therefore .1782$ Nepers per mile

Distance Miles = d	Bels	Decibels	$\left(\frac{1}{10^{.06}}\right)^d = \frac{1}{10^{.06d}} =$	$=$	P Milliwatts
0	0	0	$\left(\frac{1}{10^{.06}}\right)^0 = \frac{1}{10^{.06 \cdot 0}} = \frac{1}{1}$	= 1	1000
1	.06	.6	$\left(\frac{1}{10^{.06}}\right)^1 = \frac{1}{10^{.06 \cdot 1}} = \frac{1}{1.14015}$	= .871	871
5	.3	3	$\left(\frac{1}{10^{.06}}\right)^5 = \frac{1}{10^{.06 \cdot 5}} = \frac{1}{10^{.3}} = \frac{1}{1.995}$	= $\frac{1}{2} = .500$	500
10	.6	6	$\left(\frac{1}{10^{.06}}\right)^{10} = \frac{1}{10^{.06 \cdot 10}} = \frac{1}{10^{.6}} = \frac{1}{3.98}$	= $\frac{1}{4} = .250$	250
15	.9	9	$\left(\frac{1}{10^{.06}}\right)^{15} = \frac{1}{10^{.06 \cdot 15}} = \frac{1}{10^{.9}} = \frac{1}{7.94}$	= $\frac{1}{8} = .125$	125
20	1.2	12	$\left(\frac{1}{10^{.06}}\right)^{20} = \frac{1}{10^{.06 \cdot 20}} = \frac{1}{10^{1.2}} = \frac{1}{15.85}$	= $\frac{1}{16} = .0625$	62.5
25	1.5	15	$\left(\frac{1}{10^{.06}}\right)^{25} = \frac{1}{10^{.06 \cdot 25}} = \frac{1}{10^{1.5}} = \frac{1}{31.6}$	= $\frac{1}{32} = .03125$	31.25
30	1.8	18	$\left(\frac{1}{10^{.06}}\right)^{30} = \frac{1}{10^{.06 \cdot 30}} = \frac{1}{10^{1.8}} = \frac{1}{63.1}$	= $\frac{1}{64} = .015625$	15.625
35	2.1	21	$\left(\frac{1}{10^{.06}}\right)^{35} = \frac{1}{10^{.06 \cdot 35}} = \frac{1}{10^{2.1}} = \frac{1}{125}$	= $\frac{1}{128} = .0078125$	7.8125
40	2.4	24	$\left(\frac{1}{10^{.06}}\right)^{40} = \frac{1}{10^{.06 \cdot 40}} = \frac{1}{10^{2.4}} = \frac{1}{250}$	= $\frac{1}{256} = .00390625$	3.90625
45	2.7	27	$\left(\frac{1}{10^{.06}}\right)^{45} = \frac{1}{10^{.06 \cdot 45}} = \frac{1}{10^{2.7}} = \frac{1}{500}$	= $\frac{1}{512} = .001953125$	1.953125
50	3.0	30	$\left(\frac{1}{10^{.06}}\right)^{50} = \frac{1}{10^{.06 \cdot 50}} = \frac{1}{10^3} = \frac{1}{1000}$	= $\frac{1}{1024} = .0009765625$	.9765625

The above data were used in plotting the first of the two curves previously provided.

FIGURE 51

Original from
CORNELL UNIVERSITY

The equation is readily written $P = P_0 (10^{-.06})^d$. If we write the equation with e as a base the equation readily becomes

$$P = P_0 (e^{-.1382})^d$$

It is to be remembered that $10^{-.06}$ and $e^{-.1382}$ have the same numerical value; namely .871.

$$10^{-.06} = \frac{1}{10^{.06}} = e^{-.1382} = \frac{1}{e^{.1382}} = \frac{1}{1.14816} = .871$$

It is to be observed again that we have equation of power decay in the general terms of e^{-x} . It is arrived at in a simple, understandable, straightforward way.

LOGARITHMS OF POWER TO BASE 10

Let us now do some more background thinking and analysis.

The two accompanying charts (Figures 50, 52) each have two curves. One chart portrays geometric retrogression based on the logarithmic base 10 and the other chart portrays the same facts but is based on the logarithmic base 2.7182. From the standpoint of understanding the simple truths involved, there is no preference. However, since one chart involves e and e^{-d} , it appears more promising in connection with the objective: What are sinh and cosh good for?

The power curve itself is $P = P_0 10^{-.06d}$. Now we can do some simple maneuvering.

We can take logarithms and write an equation according to basic doctrines of logarithms

$$\log_{10} P = \log_{10} P_0 + \log_{10} 10^{-.06d}$$

It is at once observed that $\log_{10}$ of $10^{-.06d}$ is $(-.06d)$ itself.

The equation then becomes, $\log_{10} P = \log_{10} P_0 - .06d$

As is noted on chart, the .06 bels per mil circuit characteristic constant can, if we so choose, be changed to .6 decibels per mile by multiplying by 10. Bel is a name given to the logarithm to the base 10. There are 10 decibels in one bel. This ascribing a name to a pure number which is an exponent or logarithm is merely a matter of convenient language. The word bel is taken from name of Alexander Graham Bell.

As has been shown, we can write the decay equation in terms of 2.718^{-cd} as well as 10^{-kd} . If we choose to use the 2.718^{-cd} , then the constant (c) in the exponent is in nepers per mile. If we use the 10^{-kd} , the constant (k) in the exponent is in decibels per mile. The distance is " d " in both cases.

P_0 in the previous equation can be assumed to be 1000 milliwatts and the log or exponent of 1000 to base 10 is 3 bels: $10^3 = 1000$.

$$\text{Therefore } 10 \log_{10} P = 30 \text{ decibels} - .6d \text{ decibels}$$

$$\text{Decibels of } P = 30 \text{ decibels} - .6d \text{ decibels}$$

where d is the distance in miles out from starting point.

As per simple principles of analytic geometry, this equation is the equation of straight line of conventional form $Y = Y_0 - mX$.

Y_0 is the Y axis intercept namely 30 and $-m$, namely $(-.6)$, is the negative slope of the straight line. The resulting straight line has been plotted on the chart. The values of Y , namely $10 \log P$, (decibels) are given along the right hand vertical reference line.

Let us put the values in tabular form (Figure 51) by which the straight curve was plotted. The values of P come from geometric retrogression equation shown at the top of Figure 51.

Since the decay or attenuation has occurred in accordance with nature's imperious levy of power reduction as per geometric retrogression doctrine one can say; unlike the gasoline in an automobile tank which decreases in direct proportion to increasing distance (an arithmetic decay), the loss of power in a telephone circuit decreases in such a manner, so that the logarithm of the power decreases directly proportional to the increased distance. It is to be emphasized that it is not the power that decreases directly proportional to the increased distance but it is the logarithm of the power.

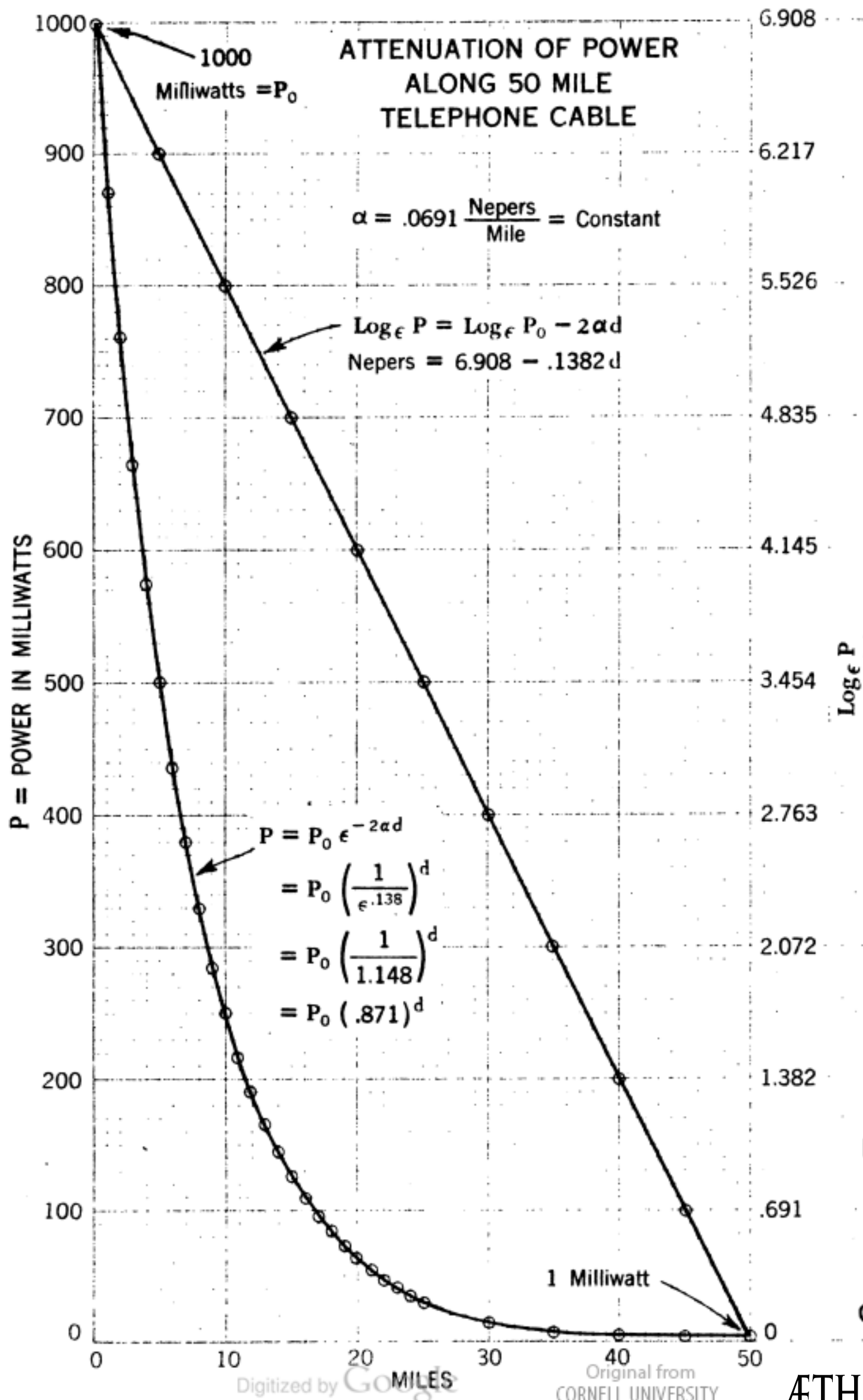


FIGURE 52

CURVE 2

TABLE 2

Some Calculations for a Telephone Cable Circuit
Using Logarithms to the Base 2.7183(e)

$$P = P_0 (.871)^d = P_0 \left(\frac{1}{1.14915} \right)^d = P_0 \left(\frac{1}{e^{.1382}} \right)^d = P_0 \left(\frac{1}{e^{2 \cdot .0691}} \right)^d = P_0 \left(\frac{1}{e^{2a}} \right)^d = P_0 \frac{1}{e^{2ad}} = P_0 e^{-2ad}$$

$$a_1 = .0691 \text{ Nepers per Mile}$$

$$2a_1 = .1382 \text{ Nepers per Mile} = \text{Power Attenuation Characteristic of Circuit}$$

$$= .06 \text{ Nels per Mile} = .6 \text{ Decibels per Mile}$$

Distance Miles - d	Nepers a - da ₁	$\left(\frac{1}{e^{2a_1}} \right)^d$	$= \frac{1}{e^{2 \cdot .0691d}}$	$= \frac{1}{e^{.1382d}}$	$= \frac{1}{e^{.1382}}$	$= \frac{1}{1.14915}$	$= .871$	P Milliwatts
0	0	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^0 = \frac{1}{e^{2 \cdot .0691 \cdot 0}} = \frac{1}{e^{0}} = \frac{1}{1} = 1$					1	1000
1	.0691	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^1 = \frac{1}{e^{2 \cdot .0691 \cdot 1}} = \frac{1}{e^{.1382}} = \frac{1}{1.14915} = .871 = .871$						871
5	.3454	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^5 = \frac{1}{e^{2 \cdot .0691 \cdot 5}} = \frac{1}{e^{.6905}} = \frac{1}{e^{.6905}} = \frac{1}{1.995} = \frac{1}{2} = .871^5$						500
10	.6908	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{10} = \frac{1}{e^{2 \cdot .0691 \cdot 10}} = \frac{1}{e^{1.382}} = \frac{1}{e^{1.382}} = \frac{1}{2.98} = \frac{1}{4} = .871^{10}$						250
15	1.0362	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{15} = \frac{1}{e^{2 \cdot .0691 \cdot 15}} = \frac{1}{e^{2.0727}} = \frac{1}{e^{2.0727}} = \frac{1}{7.94} = \frac{1}{8} = .871^{15}$						125
20	1.3820	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{20} = \frac{1}{e^{2 \cdot .0691 \cdot 20}} = \frac{1}{e^{2.764}} = \frac{1}{e^{2.764}} = \frac{1}{15.85} = \frac{1}{16} = .871^{20}$						62.7
25	1.7269	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{25} = \frac{1}{e^{2 \cdot .0691 \cdot 25}} = \frac{1}{e^{3.4547}} = \frac{1}{e^{3.4547}} = \frac{1}{31.9} = \frac{1}{32} = .871^{25}$						31.9
30	2.0723	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{30} = \frac{1}{e^{2 \cdot .0691 \cdot 30}} = \frac{1}{e^{4.1427}} = \frac{1}{e^{4.1427}} = \frac{1}{63.1} = \frac{1}{64} = .871^{30}$						15.9
35	2.4179	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{35} = \frac{1}{e^{2 \cdot .0691 \cdot 35}} = \frac{1}{e^{4.8326}} = \frac{1}{e^{4.8326}} = \frac{1}{125.9} = \frac{1}{128} = .871^{35}$						8
40	2.7631	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{40} = \frac{1}{e^{2 \cdot .0691 \cdot 40}} = \frac{1}{e^{5.5244}} = \frac{1}{e^{5.5244}} = \frac{1}{250}$						4
45	3.1085	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{45} = \frac{1}{e^{2 \cdot .0691 \cdot 45}} = \frac{1}{e^{6.2159}} = \frac{1}{e^{6.2159}} = \frac{1}{500}$						2
50	3.4539	$\left(\frac{1}{e^{2 \cdot .0691}} \right)^{50} = \frac{1}{e^{2 \cdot .0691 \cdot 50}} = \frac{1}{e^{6.9095}} = \frac{1}{e^{6.9095}} = \frac{1}{1000}$						1

The above data were used in plotting,
the second of the two curves previously
provided.

FIGURE 53

Distance	P = Milliwatts	Log _e P = Bels	10 Log ₁₀ P = Decibels	Decay Ratio
0	1000	3	30	
1	871	2.94	29.4	.871 = .871
5	500	2.7	27	1/2 = .871 ²
10	250	2.4	24	1/4 = .871 ⁴
15	125	2.1	21	1/8 = .871 ⁶
20	62.5	1.8	18	1/16 = .871 ⁸
25	31.25	1.5	15	1/32 = .871 ¹⁰
30	16.62	1.2	12	1/64 = .871 ¹²
35	8.	.9	9	1/125 = .871 ¹⁴
40	4.	.6	6	1/250 = .871 ¹⁶
45	2.	.3	3	1/500 = .871 ¹⁸
50	1.	0	0	1/1000 = .871 ²⁰

This attenuation process is a simple and direct result of the basic geometric retrogression doctrine characteristic of many natural phenomena. It is a quantitative decree laid down by nature and applied to a simple qualitative situation.

LOGARITHM OF POWER TO BASE 2.7182

The chart, Figure 52, portrays exactly the same facts of power attenuation according to geometric retrogression of Figure 50, but the quantitative considerations are based on a different scale or "yardstick" as it were.

The exponents or logarithms to the base e are called nepers. $P = P_0 e^{-2 \cdot .0691 d} = P_0 e^{-.1382 d}$ ($e^{-.1382} = 10^{-.06} = .871$)

The α (Greek letter alpha) is a constant of the circuit and as shown is the nepers per mile, .0691. Why α is multiplied by 2 will be fully and simply discussed in the following pages. Not only is the constant in the exponent or logarithm to base 2.7182 a neper per mile but also it is arbitrarily ascribed a particular notation, α . Bels per mile are, by contrast, not ascribed any special Greek letter to identify them.

As before in the case of the equation written in terms of 10 to an exponent, we can write the above equation,

$$\log_e P = \log_e P_0 + \log_e e^{-.1382 d}$$

It is to be noted that the $\log_e e^{-.1382 d}$ is $-.1382 d$ itself.

$$\text{Therefore } \log_e P = \log_e P_0 - .1382 d$$

$$\begin{aligned} \text{Nepers of } P &= \text{Nepers of } P_0 - .1382 d \text{ Nepers.} \\ &= 6.908 - .1382 d \end{aligned}$$

This is the equation of a straight line as per analytic geometry doctrine.

The Y 's or ordinates are $\log_e P$ in nepers. There need be no mystery nor misgivings about the Y axis values. The log of 1000 to the base 2.718 is 6.908; $e^{6.908} = 1000$.

Log of 500 to base is 3.454; $e^{3.454} = 500$.

The nepers or logs of P to base e , were readily obtained from Napierian log tables. They also could have been easily calculated by multiplying the logs of P to the base 10 in previous table by the factor 2.3026. The basic equation relating these logs is $\log_e P = \log_e 10 \log_{10} P = 2.3026 \log_{10} P$.

It may be a bit of fun for the reader to follow a bit of checking as for example.

The $\log_{10}$ of 1000 (0 miles) is 3; $10^3 = 1000$. The $\log_e 1000$ is 6.908, namely 3 times 2.3026. The $\log_{10}$ of 871 milliwatts (1 mile) is 2.94. The log of 871 to base e is 2.94 times 2.3026 namely 6.77. This value is readily estimated from curve.

The $\log_{10}$ of the 500 milliwatts at 5 miles is provided by the point (directly above $P = 500$) on the "logarithm to base 10 straight line" as 2.7. The log to base e of 500 on second chart is shown as 6.217 namely 2.7 times

2.3026. The $\log_{10}$ of 250 milliwatts (10 miles) is provided by the point directly above $P = 250$, on the "logarithm to base 10 straight line", as 2.4. The $\log_e$ of 250 is provided by the point directly above $P = 250$, on straight line to base e , as 5.526 namely 2.4 times 2.3026. The $\log_{10}$ of 100 milliwatts (16.5 miles) on straight line of first chart is seen to 2. The corresponding point on straight line of second chart is 4.6, namely 2 times 2.3026. It is to be noted that the scale for logs to base 10, on straight line of first chart, are .3 per 10 divisions of plotting paper. For the logs to base 2.718, on straight line of second chart, the scale is .691 for 10 divisions of plotting paper. This checks all right since $.3 \text{ times } 2.3026 = .691$.

POWER LOSS RATIO VARIES AS FREQUENCY

In the discussion in previous pages, the power loss ratio per mile for a telephone cable was assumed to be .871. Actually the loss ratio per mile in telephone circuits varies as to type of circuit and as the frequency of the alternating current involved.

The .871, $10^{-.03d}$ and $e^{-.138d}$ figures would apply only to a particular cable circuit at a particular frequency and constitute a sort of weighted average.

A particular telephone cable might have the following different attenuation or decrement constants.

Cycles/Sec	Loss per mile (db)	Loss per mile (nepers)	Loss ratio per mile
50	.3	.069	.933
500	.9	.207	.813
1000	1.3	.299	.741
5000	2.5	.576	.562

The equation (typical of all cases) for one of the above frequencies would be $P = P_0 (.933)^d = P_0 10^{-.03d} = P_0 e^{-.138d}$.

SOME FURTHER OBSERVATIONS ABOUT GEOMETRIC RETROGRESSION

It has been established as a simple and basic truth that in the particular telephone cable circuit under consideration that the power at the end of each mile was .871 or 87.1 % of the power at the beginning of the mile. This in turn means that 12.9 % of the power at the beginning of the mile was used up or lost during that particular mile. So it is in order to further examine this fact for the later distances.

During the first mile 12.9 % of 1000 watts was lost. During the second mile 12.9 % of 871 watts was lost and so on.

The following table interestingly shows quantitative values.

Distance	Existing Power = P	Ratio Lost	Amount Lost
0 miles	1000 milliwatts	12.9%	0
1 mile	.871 of 1000 = 871.	12.9%	129 milliwatts
2 miles	.871 of 871 = 758.6	12.9%	112.4 milliwatts
3 miles	.871 of 758.6 = 660.7	12.9%	97.7 milliwatts
4 miles	.871 of 660.7 = 575.5	12.9%	85.2 milliwatts
5 miles	.871 of 575.5 = 500.28	12.9%	75.2 milliwatts
			Sum = 499.5 milliwatts

These data check within the limits of accuracy of ratio used which are not figured out to several decimal places. The ratio of .871 is really .8709.

The ratio of powers (the P 's), is constant namely .871 per mile and the ratio of power lost is constant namely .129. The actual amount of losses for 5 miles add up to 500 milliwatts and the amount remaining is 500 milliwatts. This checks with the ratio of .5 for 5 miles; $(.871)^5 = .5$.

These data point up the distinguishing characteristic of natural decay as per doctrine of geometric retro-

gression. A constant proportion or ratio of power remains and a constant ratio is lost after each mile. However, a constant amount is not left nor lost. As will be noted, the amount lost decreases.

At the beginning of the circuit when the amount is large, the amount lost is large. Later when the amount is small (very small at 40 miles) the amount lost is small.

Geometrically this means that the slope of the line expressed by the notation of calculus is large when the amount is large. The $\frac{dy}{dx}$, the slope or rate of change of power with distance decreases. The ratio of power left and power lost over a given distance remains the same namely .871 and .129 respectively.

If we write the general equation in the particular form, $Y = e^{-cx}$, then $\frac{dy}{dx} = -ce^{-cx}$. This says the rate of change is negative and is numerically c times the amount of power existent at a certain point. Again the equation in the form, $Y = e^{-cx}$ is convenient because the slope or $\frac{dy}{dx}$ at a point is numerically c times Y or c times the power at that point, e^{-cx} . See Figure 43, p. 103.

SOME CONSIDERATIONS OF THE EQUATIONS

$$Y = Ae^{cx} \quad \text{and} \quad Y = Ae^{-cx}$$

We are now ready to embark on a particular analysis of the particular natural growth and natural decay equations portraying geometric progression and geometric retrogression, namely $Y = e^x$ and $Y = e^{-x}$

These equations or formulas are the same kind as the simple exponential ones studied in high school algebra.

$$Y = ar^x \quad \text{and} \quad Y = ar^{-x}$$

Also, there are the same kind of equations as

$$Y = a 10^{bx} \quad \text{and} \quad Y = a 10^{-bx}$$

The three types of exponential equations provide exactly the same answers if the constants are properly proportioned. They are founded on three different systems of logarithms to be sure but they are basically the same equation.

The particular forms $Y = Ae^{cx}$ and $Y = Ae^{-cx}$ also have certain advantages since sinh and cosh can be numerically evaluated if we know e^x and e^{-x} . It will be recalled that numerically

$$\text{Sinh } X = \frac{e^x - e^{-x}}{2} \quad \text{and} \quad \text{Cosh } X = \frac{e^x + e^{-x}}{2}$$

These hyperbolic expressions have been discussed and rationalized in the treatise on hyperbolic functions as such. There are tables available which provide their numerical values.

It is now proposed to embark on an analysis of the two particular equations which provide information as to growth and attenuation that nature seems disposed to prescribe in many cases for the quantitative aspects of changes in many natural phenomena. As stated several times before, they are the equations of geometric progression and geometric retrogression.

The equation $Y = e^x$ says fundamentally that X is the logarithm of Y to the base e . X is the so called independent variable and Y is the dependent variable.

If X is assigned different values namely, 0, .5, 1, 1.5, 2.0, 2.5, 3, then Y will be 1, 1.649, 2.718, 4.482, 7.389, 12.181 and 20.086 respectively. See tables of natural logarithms. The latter values are e^0 , $e^{.5}$, $e^{1.0}$, $e^{1.5}$, $e^{2.0}$, $e^{2.5}$ and $e^{3.0}$.

The constant ratio between the Y 's when X is made equal to 1, 2, 3, 4, etc. will be e or 2.718. The growth according to this doctrine therefore will be a geometric progression. $Y = e^x$ prescribes a particular and special geometric progression whose constant ratio between terms is e or 2.718. This is a rather large constant ratio 271.8%. Therefore, the equation is generally written $Y = e^{cx}$ in which the constant ratio is usually less than e but still greater than unity.

If the ratio in a particular case is less than unity, then the equation is $Y = e^{-x} = \left(\frac{1}{e}\right)^x = .368^x$. Then the geometric retrogression or attenuation results. The .368 is quite a sizable reducing ratio namely 36.8% so the equation is oftentimes written $y = e^{-x}$.

If the X is taken as 0, -.5, -1, -1.5, -2, -3, -4 and -5 then the geometric attenuating values of Y or $\left(\frac{1}{e}\right)^x$ are 1, .6065, .3679, .2231 and .1353, .0498, .0183 and .0067. These are evaluations of $\left(\frac{1}{e}\right)^0, \left(\frac{1}{e}\right)^{.5}, \left(\frac{1}{e}\right)^{1.0}, \left(\frac{1}{e}\right)^{1.5}, \left(\frac{1}{e}\right)^{2.0}, \left(\frac{1}{e}\right)^3, \left(\frac{1}{e}\right)^4, \left(\frac{1}{e}\right)^5 \left[\left(\frac{1}{e}\right)^5 = .368^5 = .0067 \right]$

The accompanying curves illustrate graphically this particular natural geometric growth and decay whose constant ratio is e (2.718), a rather large constant ratio for growth and decay.

THE NAPERIAN BASE AND NATURAL GROWTH AND DECAY

The curves of the preceding sketch (Fig. 52) are sometimes designated as "natural" growth and "natural" decay. It is true that many natural phenomena which involve geometric progression and attenuation can nicely and simply be considered quantitatively by invoking the equations $Y = e^x$ and $Y = e^{-x}$. It seems to the writer that the word "natural" is especially appropriate. However, some text books give the impression that $Y = e^x$ and $Y = e^{-x}$ are the only and special equations which prescribe natural growth and decay. There is the connotation that in the e there is a special significance which magically fits natural growth and decay. The decay and growth doctrines are natural but the " e " is artificial and the entire values of 2.718 or $\frac{1}{e}$ (.368) are seldom if ever encountered as the constant ratio. In other words, the equations are usually $Y_1 = e^{cx}$ and $Y_2 = e^{-Kx}$. The constants " c " and K is usually less than unity since in natural phenomena the decays and growths are quite gradual. A number of examples of natural decay are discussed in previous pages 110, 113, 114 of this treatise. Growth and decay can nicely and simply be considered quantitatively by adopting and invoking e^x and e^{-x} . But when one considers growth and decay from the standpoint of basic understanding, one does not need logarithms to the base 10 or to the base 2.718. Use of logarithms merely facilitates numerical calculation. The simple and basic doctrine of natural growth and decay is not given better mathematical expression or more accurate portrayal by the use of " e " than by the use of any one of a number of logarithmic bases.

The Napierian logarithms and the base e originally came out of the fact that in differential calculus, the handling of the differential of $\log X$ was facilitated by use of e . When Newton and Leibnitz were developing the technique of calculus and particularly the differential of $\log X$, they were confronted with a factor in the differential involving $\left(1 + \frac{1}{n}\right)^n$. By making e , the evaluation of $\left(1 + \frac{1}{n}\right)^n$ when $n = \text{infinity}$, the base of a logarithmic system, the derivative of $\log X$ to the base " e " was simple, namely $\frac{d}{dx} \log_e X = \frac{1}{X}$.

Furthermore the derivative of e^x came out simply, namely e^x and derivative of e^{-x} , came out as $-e^{-x}$. The derivative of $e^{-x} = -e^{-x}$.

The writer would go so far as to say that the Napierian logarithms and e are artificial rather than natural. But they do work and provide a fine tool for quantitative analysis of natural growth and decay problems.

The derivative $\frac{dy}{dx}$ of $\log_a u$, where " u " is a function of X involves, in the sheer process and technique of differential calculus, the expression $\left(1 + \frac{1}{n}\right)^n$. If the $\left(1 + \frac{1}{n}\right)^n$ or e when $n = \text{infinity}$ is injected into the

process then $\frac{d}{dx} \log_a u = \log_a e \frac{du}{u dx}$

Then $\frac{d}{dx} \log_e X = \frac{1}{X}$, a simple process and a simple result.

So the "e" becomes a good and convenient mathematical tool and does facilitate mathematical and numerical handling of growth and attenuation problems as it is associated numerically with hyperbolic functions. However "e" has no magic, mysterious, extraordinary, transcendental, imaginary or subtle meaning. It is a number of a particular numerical value but it makes no contribution whatever to understanding and insight.

GEOMETRIC GROWTH AND DECAY: A BIT OF REVIEW

It will be recalled that there are two significant characteristics of geometric growth and attenuation. The first characteristic is that the ratio between any two successive terms in the growth or decay is a constant. The growth and decay curves written in at least three forms comply with this specification.

The second characteristic is that the instantaneous rate of change of the two factors involved is numerically equal to the magnitude of the decaying or growing factor dependent on time or distance at the instant of time or particular distance involved.

If we consider the growth and decay in terms of $Y = e^x$ and $Y = e^{-x}$, then the derivative $\frac{dy}{dx}$, which is the instantaneous value of rate of change, of $Y = e^x$ is e^x of $Y = e^{-x}$ is $-e^{-x}$. The mathematics supports the validity of the second characteristic and provides a numerical value for the slope of the curve or graph portraying the growth or decay.

SLOPE OF CURVES AND DERIVATIVES

The derivative of calculus as viewed from the standpoint of curves is basically the slope of the curves. So in the curves of e^x and e^{-x} in previous page it will be noted that the Y axis not only portrays the value of $Y = e^x$ but also the slope of the curve (e^x) numerically equal to Y or e^x at each point along the curve. See Figure 43, page 103.

This says that when Y or e^x is large, the slope of the curve is large and is positive. When $Y = e^{-x}$, and Y is also small then e^{-x} is small. The slope is numerically e^{-x} and negative and also small. Inspection of the curves substantiates the ideas provided by the mathematics and the validity and significance of the second characteristic of geometric growth and attenuation.

At point where curve $Y = e^x$ crosses Y axis, the X is zero and the $Y = e^0 = 1$. The $\frac{dy}{dx}(e^0)$ at this point is also e^0 and therefore equal to $+1$. By inspection one observes that the slope of the curve at this point is 45 degrees or $+1$.

At this same point on the $Y = e^{-x}$ curve when $X = 0$, Y is 1 and the slope is negative and equal to -1 and again corresponds to a slope of -45 degrees.

The fact that the value of Y , at X zero, is 1 and that the slope of the curve or $\frac{dy}{dx}$ is also 1, illustrates the point made that if $Y = e^x$ is used for the equation of growth, the rate of change at any point along the curve is numerically equal to the value of the function at that point. If Y is large corresponding to a large value of X , the rate of change $\frac{dy}{dx}$ will be large. This point has been made several times previously but it merits emphasis.

TELEPHONE POWER CURVE

The attenuation of power along a telephone circuit will now be considered in more complete fashion. The power versus distance curves are shown in charts, pages 125, 128, Figures 50, 52.

The equation of the curve is

$$P = P_0 (.871)^d = P_0 10^{-.06d} = P_0 e^{-2ad} = P_0 e^{-.1382d}$$

1. P_0 is the original amount of power at input of circuit in milliwatts.
2. P is the power at some distance along the circuit.

3. e is the Naperian logarithmic base, 2.7182.
4. α is a constant in nepers per mile characteristic of the type of circuit under consideration.
5. " d " is the distance in miles from original point at which distance the power has attenuated from P_0 to P .

It is now proposed to discuss and argue about this equation from many and various fundamental points of view.

In the first place, it is to be noted that the equation is written in three exponential forms of a type which specifies the so called natural attenuation or decay fashion.

In the particular cable circuit in question, the constant ratio between the powers (milliwatts) at two points 5 miles apart is 2 or $\frac{1}{2}$ depending on which way one travels along the circuit. The logarithms, to the base 10 of two successive powers whose ratio is 2, varies by .3 bels or 3 decibels. The logarithm of 2 to the base 10 is .3. It is to be noted here that in general one would be disposed to say that the log of 2 to the base 10 is just .3 and not .3 bels. But the phrase 3 bels would be valid.

It is now proposed to find out in a step by step fashion, the meaning of α and the reason for using the 2 in the exponent and to relate the attenuation in decibels to equivalencies in natural logarithms as per the basic general equation $Y = e^{-x}$.

The term "nepers" will again be defined, discussed, and used. The accepted fact and principle that power in any electrical circuit depends on the square of the current and on the square of the voltage (other factors being equal) will be invoked to help in understanding.

NEPERS DEFINED

As stated in the preceding pages no name or designation in general has been given to a particular amount of constant ratio between the successive terms of a geometric progression or retrogression.

In the quantitative techniques which engineers and scientists always consider their problems, units or amounts of the various factors obviously must be established.

As the electrical power attenuates along a circuit, the current and voltage also decrease. It has been generally agreed that if the ratio between the currents or the voltages at two designated points on the circuit is e or 2.718, this particular amount of geometric ratio is designated as one neper. The neper itself is really the exponent or logarithm of the ratio 2.7182.

$$\text{Let ratio of two currents} = \frac{I_1}{I_2} = e^1 = 2.7182$$

$$\text{Log}_e \frac{I_1}{I_2} = \log_e e^1 = 1 = 1 \text{ neper}$$

The ratio of e^1 of 2.7182 may be identified by the exponent of e^1 or 1 neper.

If the ratio that exists just happens to be 2.718, then the ratio is e^1 where the 1 in the exponent is the 1 neper. If the ratio were indicated by 2 nepers, the value of the ratio in ordinary notation would be e^2 or 7.389. If there were a decay of 1 neper between the two currents or voltages, then the sheer numerical value of the ratio would be e^{-1} or .3679 which is $\frac{1}{2.718}$.

If the decay ratio between the two currents or voltages were 2 nepers the sheer numerical value of the decay would be e^{-2} or .1353 = $\frac{1}{e^2} = \left(\frac{1}{e}\right)^2 = \frac{1}{7.4} = (.368)^2$. Therefore, the neper denotes or specifies quite a sizable ratio as ratios go.

It indicates a 271 % gain or 37 % loss. The $e^{-1} = \frac{1}{e} = .3679$.

It is now to be stated that losses and gains in ratio of currents and voltages are now expressed in terms of α , which is a symbol given to the constant in the exponent of the Naperian logarithmic base which serves at a basis of ratios. The α is the alpha of the Greek alphabet. It is the Greek equivalent of our English "a".

If we know the ratio of growth or attenuation of current or voltage in terms of nepers we can readily find the ratio in sheer numerical value by looking up the value e^x and e^{-x} in a book of tables. The constant in the

exponent X is given in nepers per mile is usually designated as α . It is a specific notation for a specific problem. The general X has been replaced by αd . Really, it is not correct to say, "ratio in nepers". The nepers per mile designated by α are the constants in exponents of " e " which in turn enable ratios to be found, namely values of e^x and e^{-x} in a book of tables. The numerical value of X will be αd .

POWER RATIO

It is to be observed that nepers specifically and solely arbitrarily pertain to ratios of currents and voltages.

Since a power ratio would involve the square of the current or voltage ratio, one would need to square the ratio as indicated in nepers in the desire to determine power ratio. The squaring of the ratio requires that the exponent be multiplied by 2. $(e^\alpha)^2 = e^{2\alpha}$.

So if the current or voltage were attenuated by a ratio identified as 1 neper or .3679 (e^{-1}) in ratio, the power involved would be attenuated by a ratio identified as 2 nepers or e^{-2} ; $(.368)^2 = .1353$.

If the nepers (current or voltage ratio), were .35 ($e^{-.35}$), the power ratio would be $e^{-.7}$ or $e^{-2 \cdot .35}$ or $e^{-.7}$ which in sheer numerical value is $\frac{1}{2}$. The sheer numerical value of current or voltage ratio corresponding to an α of .35 nepers or $e^{-.35}$ would be .7047. The .7047 squared also provides the $\frac{1}{2}$ for power ratio;

$$e^{-.7} = (e^{-.35})^2 = (e^{-.35})^2 = .7047^2 = e^{-.7} = .5$$

BELS AND DECIBELS

The nepers are the exponents or logarithms to the base e and specifically provide the numerical value of ratio of two currents or two voltages by evaluating the $e^{-\alpha}$.

The decibel, as its name implies is $\frac{1}{10}$ of a bel. One bel equals 10 decibels. The bel is the name given to the logarithm or exponent of the base 10 which will provide the value of a ratio numerically equal to 10; $10^1 = 10$.

A ratio of growth or increase of 10 decibels or 1 bel between two factors would mean the ratio is 10^1 or 10. A ratio of decay of 10 decibels or 1 bel would be 10^{-1} , $\frac{1}{10}$ or .1 in sheer numerical value.

It is sometimes stated, rather erroneously it seems to the writer, that bels and decibels apply and are predicated particularly on power increase and power decrease. Actually current and voltage ratios could be given in bels and decibels as well as nepers. It just happens that it arbitrarily has been agreed upon that nepers apply particularly to current or voltage ratios in considering problems of their geometric increase and decrease. If attention is given to the power ratio then it is the practice and general agreement to consider this power ratio in bels or decibels. This agreement is an entirely arbitrary one. Bel and decibel are not set up to apply to power ratios any more than to many other factors which increase and decrease. Since nepers are specifically assigned for consideration of ratios of currents and voltages per se then power ratios may be evaluated in terms of bels and decibels. The neper denoted by α to indicate current and voltage ratio exponents must be multiplied by 2 to get ratio of power since power involves the square of current and voltage. The power ratio can then be expressed in neper equivalencies for which equivalencies the decibel is arbitrarily used.

A general and simple fact now merits a restatement, nepers and bels are exponents or logarithms. In the case of nepers, the base is e (2.718). In the case of bels, the base is 10. In both cases, the numbers corresponding to the base raised to some power are the constant ratios between any two successive terms of a geometric progression or retrogression.

Nepers are arbitrarily and specifically designed to apply to current and voltage ratios. Bel and decibel are exponents to the base 10 and can be used in considering ratios of any quantity among which is power in watts or milliwatts. They could very well be used for ratios of current and voltage if nepers were not reserved as it were for this purpose.

THE POWER DECAY CURVE

The power curve as given in Figure 52 on previous page (128) was $P = P_0 e^{-2\alpha d}$.

The α in this equation is the nepers per mile which characterized the current ratio per mile along the circuit.

The equations for current attenuation might be

$$I = I_0 r^d = I_0 10^{-kd} = I_0 e^{-\alpha d}$$

The " r " is less than unity (1), $\frac{1}{10^k} = \frac{1}{e^\alpha} = r$

The number 2 must be used therefore in the above power equation to take care of the quantitative fact that power involves current squared. If the P is power, then the nepers (the α), must be multiplied by 2 to arrive at the proper exponent in the equation for power attenuation or decay.

INTERCHANGE AND EQUIVALENCY OF NEPERS AND DECIBELS

It is now in order to discuss how ratios indicated in nepers and decibels are interrelated. The procedure is a simple and straightforward one. Several specific examples will be used to develop the interrelationship. Finally, a general expression for the relationship will be given.

Assume, as a beginning example, that the ratio between the two currents is 2.718 or e^1 . The nepers, indicated by the Greek letter α , will be one (1). The power involved will be e^2 and numerically the ratio would be 7.389. The logarithmic base 2.718 must be raised to the 2nd power to get a numerical value of 7.389 for the ratio. The nepers would be 2.

Now the simple question that suggests itself is, to what power must the logarithmic base 10 be raised to get a numerical value of 7.389 for the ratio. The answer is easily found in the $\log_{10}$ table as .8686. The bels would be .8686.

$$e^2 = 10^{.8686} = 7.389$$

One can readily change the .8686 bels to 8.686 decibels. So 1 neper in current ratio represents or is equivalent to 8.686 decibels in power ratio.

As a second simple approach to find the equivalency of nepers and decibels, we can use the following straightforward argument.

Assume $e^{2\alpha} = e^{2 \cdot 1.1513} = 2.3026 = 10 = 10^1$.

The ratio would be 10 as the result of 2.3026 nepers (1 bel). The 2.3026 nepers is 2α to achieve a power ratio of 10. The α would be 1.1513. The number of decibels is 10. Therefore 10 decibels in power ratio are equivalent to 1.1513 nepers in current ratio. So 1 neper in current ratio represents or is equivalent to 8.686 decibels in power ratio, namely 10 divided by 1.1513.

SOME MORE CHECKING IN TERMS OF NUMERICAL RATIOS OF ATTENUATION

If two currents has a ratio of .0691 nepers per mile the sheer numerical value would be $e^{-.0691} = \frac{1}{e^{.0691}} = \frac{1}{1.0715} = .933$.

For power ratio per mile, one would have tables

$$e^{-2 \cdot .0691} = e^{-.1382} = \frac{1}{e^{.1382}} = \frac{1}{1.1482} = .8709 = (.933)^2$$

$$P = P_0 e^{-2 \cdot .0691} = P_0 .871 \text{ per mile}$$

A .6 decibel loss per mile which would be .06 bel loss per mile

$$P = P_0 10^{-.06} = P_0 .871 \text{ per mile}$$

$$10^{-.06} = \frac{1}{10^{.06}} = \frac{1}{1.1482} = .871 \text{ per mile.}$$

From the standpoint of the above example of sheer ratio in geometric retrogression, .0691 nepers represents or is equivalent to .6 decibels.

ANOTHER EXAMPLE

If two currents had a ratio indicated by 1 neper, the sheer numerical value of the ratio would be e^{-1} or $\frac{1}{2.7182}$ or .3679.

For power ratio, involving square of current, one would have e^{-2} or .1353 in sheer numerical value. $\frac{1}{e^2} = \frac{1}{7.389} = .1353 = (.3679)^2$

$$P = P_0.1353$$

It has been stated above that 1 neper in current decrease which is really 2 nepers in power decrease was represented by 8.686 decibels or .8686 bels.

Now $10^{-.8686}$ turns out to be $\frac{1}{7.389}$ which in turn is .1353 in sheer numerical value for the ratio. So with an 8.686 db loss

$$P = P_0.1353 = P_0 10^{-.8686} = P_0 e^{-2}$$

Again then, 1 neper is equivalent to 8.686 decibels (.8686 bels).

ONE DECIBEL EQUIVALENT TO .115 NEPERS

A one decibel loss in power would represent a .794 for sheer numerical value of ratio decrease. To achieve a ratio of .794, one needs a certain value for nepers to get the proper numerical value of exponent for "e" so that $e^{-2\alpha} = .794 = \frac{1}{10^{.1}} = \frac{1}{1.259} = 10^{-.1} = (.8913)^2$

It is readily found that the α must be .115. Therefore $e^{-.115}$ providing a current ratio of .8913 calls for $e^{-.230}$ to indicate power ratio of .794 or $(.8913)^2$.

From the book of natural log tables, $e^{-.230}$ provides a sheer numerical value of ratio of .794. This is the same value for ratio is given above for 1 decibel; $10^{-.1}$. Therefore 1 decibel represents or is equivalent to .115 nepers.

In other words $10^{-.1}$ provides the same numerical value (.794) for the magnitude of ratio of power as $e^{-.230}$ or $e^{-2 \cdot .115}$.

If simple arguments involving ratios determined the equivalency that 1 neper equals 8.686 decibels, then the simple reciprocal fact would be that 1 decibel equals .115 nepers.

However it was thought helpful to work out the latter equivalency by using the actual ratio argument to provide insight.

Finally one can formally not only write down but also readily understand from point of view of equivalency of geometric retrogression numerical values for the ratios involved, the following:

- A. 1 neper in current decrease represents or is equivalent to 8.686 decibels in power decrease.
- B. 1 decibel represents or is equivalent to .115 nepers. Again a neper is the logarithm to the base e of a number. A decibel is .1 of the logarithm of a number to the base 10.

The number involved in a ratio of decrease, of decay, or attenuation. The ratio involved in every case is the ratio of any two successive terms in a geometric retrogression.

The gist of the problem at hand relates to loss in power as an electromagnetic wave impulse carries a communication signal along a pair of wires or through an electrical network of some sort.

SUMMARY

$$\text{Nepers} = \log_e \frac{I_1}{I_2} = \log \frac{E_1}{E_2} \text{ where } I\text{'s are currents and } E\text{'s are voltages.}$$

Power ratio in number of decibels = N_{db}

$$\begin{aligned} N_{db} &= 10 \log_{10} \left(\frac{I_1}{I_2} \right)^2 = 20 \log_{10} \frac{I_1}{I_2} \\ &= 20 \log_{10} \frac{E_1}{E_2} \end{aligned}$$

$$1 \text{ decibel} = .115 \text{ nepers}$$

$$\text{Nepers} = 2.3026 \log_{10} \frac{I_1}{I_2} = \frac{2.3}{20} N_{db} = .115 N_{db}$$

$$1 \text{ Neper} = 8.686 \text{ decibels}$$

$$\log_e x = 2.3026 \log_{10} x$$

$$\log_{10} x = .4343 \log_e x$$

All natural decay or geometric retrogression phenomena can be quantitatively expressed by an equation stated in at least three exponential forms which are intrinsically identical.

The first form is $Y = Y_0 r^x$ where r is the ratio between any two successive terms and is numerically less than unity (1).

$$\text{The second form is } Y = Y_0 10^{-kx} = Y_0 \left(\frac{1}{10^k} \right)^x$$

$$\text{The third form is } Y = Y_0 e^{-cx} = Y_0 \left(\frac{1}{e^c} \right)^x$$

The " k " and " c " are logarithmic decrement constants.

The first form of the equation expresses the basic truth in a simple manner. The latter two forms of the equation really provide directions as it were for numerically evaluating Y which can be conveniently done by using either common (Briggs) or Napierian logarithms using the base 10 and base " e " (2.718) respectively.

All three equations are exponential equations. The latter two may seem to appear formidable and may be a bit frightening. Actually they are not; they merely indicate a convenient procedure to get numerical answers by using logarithms rather than a lot of plain arithmetical multiplication.

Review

CHAPTER 9

Resistance, Inductance, Capacitance

Oscillating Circuits

BASIC IDEAS, PRINCIPLES AND DEFINITIONS

The circular functions (sine and cosine), and the hyperbolic functions have been discussed as such so far. These functions are good mathematical tools with which one can work on all engineering problems in general and on electrical problems in particular. These tools are closely allied with the practical problem of energy relations in electrical systems. The energy relations must be examined and studied because efficiency is a major factor in all engineering. The word efficiency is used in its literal sense and relates to the ratio of output to input of power and energy. It is inevitably true that some factors like voltage, and electric current will decrease and peter out. The thinning out of such factors is called attenuation. This latter word may seem to be a \$64 word, but it is really the appropriate word. It conveys the simple and fundamental idea of petering out.

Attenuation problems bring one face to face in a natural and simple way with exponentials and logarithms. Attenuation in electrical systems call for some analysis of electrical circuits. Since electrical circuits of all kinds are made up of three basic circuit elements, resistance, inductance and capacitance, it is in order to consider these three elements as such.

After "wrapping" up this fundamental aspect of the overall problem, it will be in order to examine attenuation, exponentials and the utility and application of hyperbolic functions in electrical communication.

The basic purpose of electromagnetic systems, usually called electrical systems, is transmission of energy which is in the form of waves. This general statement applies to both so called electrical power systems and electrical communication systems. The waves may be guided by a variety of wire systems or may go on their own as do "radio waves". The electromagnetic waves are neither animal, vegetable or mineral; they are neither organic nor inorganic.

Mechanical systems involving belts, compressed air, water and projectiles also transmit energy. In these latter mechanical systems, a ponderable physical something is involved. They are positively not wave systems.

The two forms of energy in mechanical systems are potential energy (energy of position) and kinetic energy (energy of motion).

In electromagnetic systems, the energy may be embodied in an electric field on one hand (potential energy by comparison) and in a magnetic field on the other hand (kinetic energy by comparison).

The common denominator of energy of all forms in all types of systems is heat. Heat is the sink or sump into which all forms of energy including waves finally may be resolved. Of course, heat may be the energy desired but in many instances mechanical energy, light energy, or sound energy is the form of energy desired.

Throughout these pages, ideas, principles, and equations will be discussed from a basic and fundamental point of view. Ideas and principles will be stated in different fashions, and from different points of view; there will be repetitions, summaries and even summaries of summaries.

It takes a long time and several readings for ideas to sink in. The process of thinking and understanding depends upon lucid and explicit exposition on the one hand and a lot of meditation and study on the other hand. Good exposition calls for repetitions; not just rote repetitions but repetitions with a slightly different flavor and tang.

The teaching method aimed at in this book is that so deftly used by the composer of music. The composer uses a pretty melody or theme over and over again. One single musical instrument plays the melody. Several other single instruments repeat it. Then several instruments take up the melody in obligato fashion. A different

group of instruments with different "voices" follows at a different pitch and quality of tone. Then there be more repetitions and at the end is a finale by the entire orchestra. This musical pattern is called a fugue. The fugue technique is a good technique in learning and understanding; in acquiring "know how" and "know how much" of the basic electrical truths, facts, principles and doctrines involved.

The musicians' fugue technique will be used as a pattern worthy of emulation in the presentation of several ideas and principles all of which are essentially simple. The ensemble and finale are engaging and of practical utility in many electrical endeavors.

RESISTANCE

Resistance is that property of physical things which translates electromagnetic energy of a battery or generator into heat energy. The translation seems to be instantaneous in the case of resistance. The heat may be stored up in the mechanical or thermal system or it may be immediately radiated. It may warm something or may go down the drain, as it were, and is not recoverable. Whether the source of energy is a D. C. electrical system or an A. C. electrical system, the resistance instantaneously translates the supplied energy to heat. The heat is not stored up in the electrical systems in any wise; is immediately divorced from the electrical system; is not electrically recoverable, and may go down the drain. The heat may be useful but it is irrevocably lost as far as the electrical system itself is concerned.

A main point to be realized is the fact that the resistance of a physical thing is not dependent on frequency of the electrical energy source whether it be A. C. or D. C. A direct current may be thought of as a current of zero frequency.

Resistance is resistance; and that's that; its opposition to flow of electric current is given according to Ohm's Law, $I = \frac{E}{R}$. The energy involved is given by formula EIt or I^2Rt . At all times and at every instant the time however small, its opposition to flow of current is R , regardless of frequency. The R for a certain physical thing (wire etc.) is constant. Of course, a temperature change varies the value of R , but at a given temperature, R is constant. Resistance cannot store up heat energy in any type of an electromagnetic system so as to be recovered at some later time.

INDUCTANCE

Inductance is that property or quality of physical systems which permits of translation of electrical energy supplied by a battery or a generator into the energy of a magnetic field. See photograph, Figure 54. Inductance permits some of the energy of source to be stored in the magnetic field surrounding the wires of a coil and in the magnetic core within the coil, if the coil has an iron core. The first sentence above may be thought of as a definition of the generic property, inductance. Inductance is an innate property of a physical object which involves energy storage in the form of magnetic field. This stored energy is not lost, is not divorced from the electrical system. It is recoverable; and it can subsequently be gotten back. An inductance is a kind of energy reservoir.

The amount of energy stored in an inductance, actually embodied in the magnetic field around the single wire, an air core coil or a magnetic core coil, is dependent on two factors. The first factor is the physical inductor itself; the number of turns of the coil, the geometrical pattern of the coil and the kind of core; iron or air; these combined factors are denoted by L . The second factor governing the amount of stored energy is the amount of current in amperes guided by wires of the coil at a particular instant of time.

The formula for this stored energy is $\frac{1}{2}LI^2$.

This formula reminds one of heat energy, I^2RT in electrical systems and kinetic energy of a moving object, $\frac{1}{2}MV^2$. Inductance, L , is a factor whose magnitude depends entirely on the geometric arrangement of the coil and the nature of its core; iron or air. The physical dimensions and geometry of the coil and the properties of its magnetic core determine the amount of inductance; the number of henries. The number of henries of inductance does not depend on the current guided by its windings or the voltage, D. C. or A. C. of the source supplying the energy. The energy it stores depends on its own, L , as one factor and I^2 (current squared) as a

second factor. L itself is no more the cause of the energy storage than the current. Both are needed to store energy. L is not a cause and stored energy an effect. Both L and I are needed to achieve a result, a reservoir of energy.

WHAT IS A HENRY OF INDUCTANCE?

At this stage of the discussion, the reader may likely ask a very pertinent question. His question is, "How can one find the amount of inductance an inductor has, how can one know what a henry is?"

PUSHING A WHEELBARROW

The informant invokes a very valid analogy and argues with the questioner as follows.

"Assume I am pushing a wheelbarrow. The wheelbarrow may be empty or loaded with sand or lead.

I start to walk, pushing the empty wheelbarrow ahead of me. I start up and I increase my walking speed; going faster and faster; pushing all the time. I have a "pick up" of speed, and acceleration. What does the wheelbarrow itself have to say about the situation? Since I push and walk faster and faster, the wheelbarrow says, "I do not choose to move". It reacts and pushes back. The wheelbarrow, not only at the very start but also all the time later, reacts against being made to go faster and faster. We say it has inertia; a sort of lethargy; a sort of laziness. Nevertheless I get up some speed since I keep pushing. The empty wheelbarrow gets speed in spite of itself and therefore gets some energy of motion, some kinetic energy. I have expended some energy and put some of it into the wheelbarrow and into my own body. Some of my expended energy is used up in heat due to wind resistance and friction of wheel on pavement. There is constant frictional resistance (mechanical ohms) against which I must push as well as making the barrow gain in speed and thereby acquire more and more kinetic energy. This frictional or resistance energy is lost in heat. It is gone, it is out, it is not stored up. The other portion of my expended energy is stored in my own moving body and in the moving wheelbarrow. The wheelbarrow and I have kinetic energy; we have momentum (MV) to be sure but more significantly, we have energy.

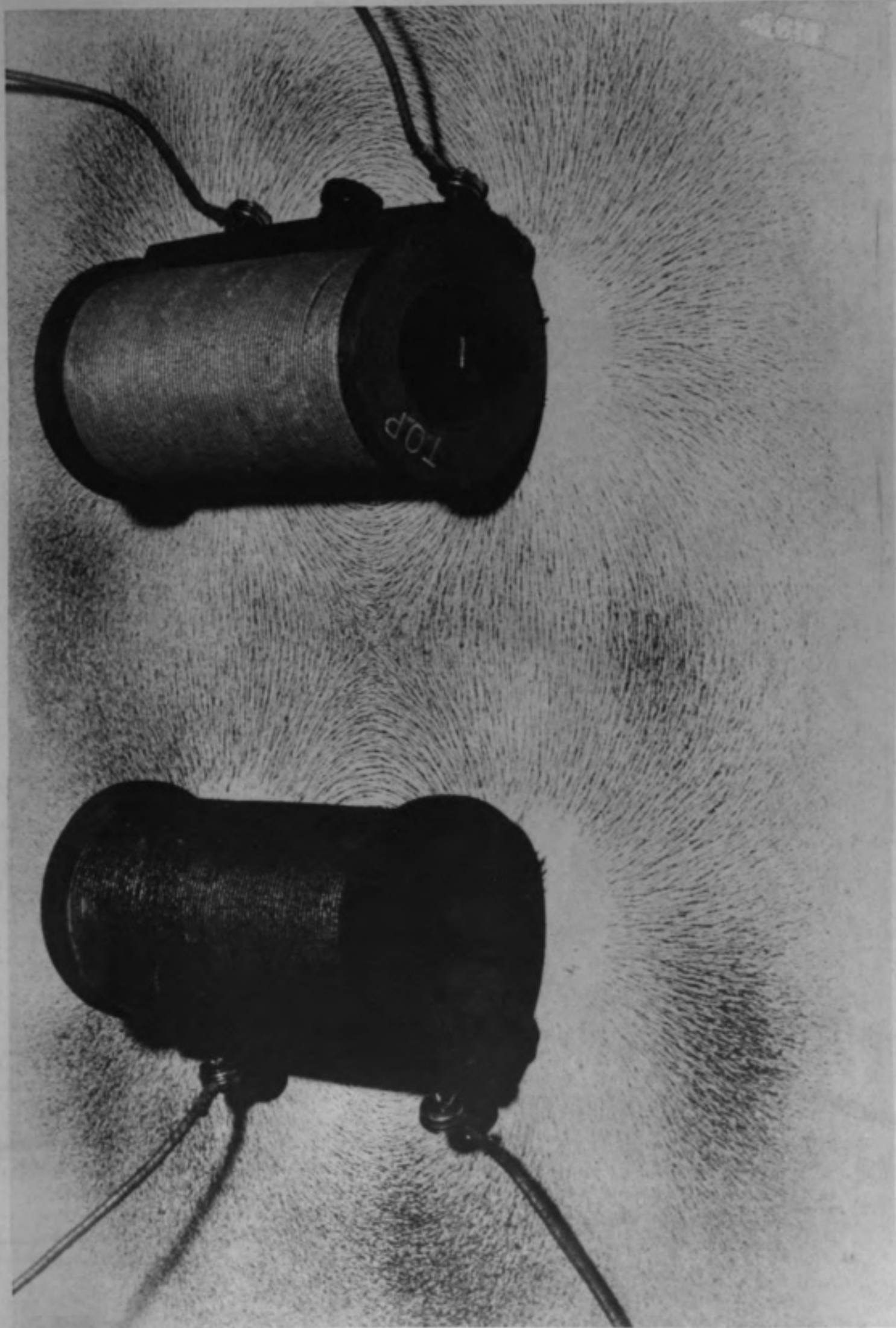
"Assume I start again from rest and with the same push, with the same force by my arms and legs, but with the barrow loaded now with sand rather than empty barrow as in the first case. I acquired more pick up in speed and had more acceleration when I pushed the empty barrow than when I push the sand-loaded barrow, provided of course as I said, I exert the same push in both cases. If I had a wheelbarrow piled high with lead blocks, I would not acquire as much acceleration with my same constant push. I have the same "voltage" or electromotive force in the three cases, i.e. I have a certain constant maximum push that I can exert."

REACTIVE FORCE PLUS RESISTANCE FORCE IS ALWAYS EQUAL TO APPLIED FORCE

In all three cases with my assumed constant push, I would get the same reactive force or opposition at all instants of time on the part of the wheelbarrow regardless of whether it was empty, loaded with sand, or loaded with lead. Each wheelbarrow reacts due to its inertia, its mass, and pushes back at all times with the "counter electromotive" force equal to the force I exert. However, even if my force and the reactive forces are equal in the three cases, my constant force gives more acceleration to the empty wheelbarrow than it gives to the sand loaded wheelbarrow. Likewise I can give more acceleration, or gain in speed, to the sand loaded wheelbarrow than the lead loaded wheelbarrow. Acceleration is change in velocity per unit time; it is expressed in feet per second per second or miles per hour per second.

In every case we say that the force involved is equal to mass times acceleration. If force is constant, acceleration (change in velocity) varies inversely as the mass. Of course I must not forget the force of resistance. Therefore to make the story complete one must really say:

"At all times and in each of the three cases my push is equaled by the wasteful force of friction plus the reactive force of inertia." My "electromotive" force is constant and the two opposing forces added together are always equal to it at every instant of time. This is a statement of a third fundamental law in electrical circuits; it is Kirchhoff's Law.



MAGNETIC FIELD SURROUNDING TWO ELECTROMAGNETS
IS PORTRAYED BY ALIGNMENT PATTERN OF IRON FILINGS

FIGURE 54

ENERGY

I have a lot of energy and a lot of "pep" to back up my push so even with the lead-loaded wheelbarrow, I manage to increase speed and get to walk faster and faster and finally running up to some maximum speed; some maximum current as an electrical analogy.

This analogy between electrical current and velocity is a good one and a valid one. Electric current equals electric charge times velocity; $I = qV = \text{charge per second; coulombs per second}$.

An ampere is a coulomb per second; a transference of electric charge per second. If one is concerned with (an increasing or decreasing) current changing with time (amperes per second) one really is justified in using the term, coulombs per second per second. This latter is a sort of electrical charge acceleration. Of course one frequently meets up with a decreasing current, a deceleration, amperes per second, or coulombs per second per second.

My maximum speed of pushing the wheelbarrow is not the absolute maximum speed I could run because the frictional resistance force finally becomes equal to my maximum constant push and I just cannot go faster. I could run faster with no wheelbarrow at all but I would be stopped in gaining speed by the air's frictional resisting force. Electrically speaking, the RI voltage "drop" finally equals my electromotive force. Obviously, it would take longer time to get the wheelbarrow full of lead rolling up to my maximum speed than when I pushed the sand-loaded wheelbarrow and also than when I pushed the empty wheelbarrow. But in all cases, I cannot run faster than the resistance permits. My final current is the same in all cases. The time factor of the lead-loaded barrow would be greatest of the three cases; it would take me a longer time to get the lead-loaded barrow up to maximum speed (current) than for the sand-loaded or empty barrow. Of course, the frictional resistance might be different in the three cases and this would alter my final speed.

My increase in velocity, my pick up, my acceleration would be greater when I pushed the empty wheelbarrow and least with the lead wheelbarrow. Since the final maximum speed (maximum electrical current by analogy) in each case is assumed to be the same on account of a fixed resistance force there would be less stored kinetic energy in the empty wheelbarrow than in the lead-loaded wheelbarrow. Myself, as the source of energy, I would deliver and expend less energy to the empty barrow than to the lead-loaded wheelbarrow. The empty barrow would have stored up less kinetic energy, $\frac{1}{2}MV^2$ than the lead-loaded barrow, $\frac{1}{2}MV^2$ since its mass (M) is smaller. It should be borne in mind that I work and put in energy for a longer time on the lead loaded barrow than I did the empty barrow. The time of getting the empty barrow up to maximum speed (maximum current) was less than in the case of the lead loaded barrow.

In all cases, I ran the same final maximum speed (current). The final wheelbarrow speed in the three cases was the same. Electrically speaking, the final mechanical "current" was the same in all three cases. I could have run faster in all cases but it was the constant frictional resistance that acted like a brake and stopped me from running faster. I always had more energy and pep available but I had only a given maximum push or voltage. When the frictional resistance provided a back force equal to my push, it was not possible for me to go faster. The RI drop was finally equal to my voltage; push.

I did expend different amounts of energy with the same velocity (current) in the three cases; getting most kinetic energy stored in the lead loaded barrow; $\frac{1}{2}MV^2$.

ELECTRICAL TERMS

Using electrical terms to describe above case, I would say; I have a battery of constant voltage and 3 circuits with 3 different inductances but with the same resistance. The same final and limiting current in amperes (final velocity) is achieved according to Ohm's Law. The empty barrow with a small mass was like a small inductance; the sand loaded barrow was like a larger inductance and the lead loaded barrow was like a still larger inductance. The electrical resistance was equal in all three cases. It is assumed that the frictional resistance was the same too. This is not a valid assumption but one can assume it to be valid. Of course the factor involved for the barrows is mass not inductance. There is energy of three different amounts with the 3 L 's but a constant current I . See Chart 1A on page 203.

Current can be thought of as a velocity. L (inductance) is analogous to mass. One kinetic energy is $\frac{1}{2}MV^2$; mechanical stored energy. The other "kinetic" energy is $\frac{1}{2}LI^2$; magnetic field stored energy.

It is in order to repeat and reemphasize several ideas in the above argument.

It was assumed I had available a given constant muscular force; that is a constant voltage to give current (velocity). The battery sought to send current through the inductance coil and the resistance.

The inductive reactance voltage (Ls) or $L \frac{di}{dt}$ and the resistance voltage drop (Ri) are at every instant of time exactly equal in combination to the battery voltage (E). The wheelbarrow pushes back only at the instants of time when the speed is increasing. The inductive reactance voltage exists only when the current is increasing. Mass reactance and magnetic reactance play a part only when feet per second per second and coulombs per second per second (amperes per second) $\frac{di}{dt}$ are involved. I, myself, was putting more kinetic energy into the wheelbarrow, all the time I, myself, could gain speed. More energy is being stored in the magnetic field only when the current increases. This energy has been supplied by the battery. An electromagnetic field energy reservoir has been present.

Finally, when the wind resistance and the frictional resistance of wheel on pavement becomes equal to my given force, I, myself, cannot gain speed; I, myself, cannot accelerate any more so the barrow stop pushing back, stop reacting and I have reached some final velocity (final current).

So when the sheer ohmic resistance, in the connecting wires and in the wires of the inductance coil, are sufficient according to Ohm's Law collectively to block any increase in current, no more energy can be stored in the magnetic field of the inductance. The current stops increasing and so the inductance stops reacting with a back voltage.

Some of the energy is irrevocably lost in resistance but energy has been stored in the magnetic field surrounding the inductance, $\frac{1}{2}LI^2$.

So in the end I keep pushing but not gaining speed; my velocity (current) remains constant. All of my expended energy is now lost in friction and no more kinetic energy is stored in kinetic energy of the barrow. The barrow no longer reacts. It stops pushing back. It's kinetic energy is a maximum.

The battery keeps pushing and sending the constant final current into the circuit; it continues to give out energy but all of its energy is now lost in heat of friction and no more energy can be stored in the magnetic field, $\frac{1}{2}LI^2$. The inductance has stopped presenting a reactive or back voltage because the current has stopped increasing.

CHARTS AND CURVES

If I plotted my velocity (current) during the time I was pushing the three barrows, I would get curves exactly like those for three different inductances portrayed in Chart 1A, page 203. The gradually rising curves show not only how my velocity increases with different masses but also how the current increases with different inductances, i.e. different L 's. My final velocity (final current) is seen to be the same in the three different cases. The longer times involved necessary to reach the final state of affairs are also portrayed in the three curves of Chart 1A, page 203.

The state of affairs when velocity and current are changing can be very appropriately called the transient state. The final state when velocity and current are not changing can be called very appropriately the steady state.

The lower left corner of Chart 1, page 279 is called the "origin". It represents the instant of time at which "things" begin to happen. It represents the instant when "time" was zero and electric current was zero; when time is zero, the velocity (mechanical current) of wheelbarrow was also zero.

With the empty wheelbarrow and my constant push (voltage), I could give a bigger change in velocity (more acceleration) at the very beginning than with the loaded barrows. So with the small inductance, the battery could give a big initial time rate of current; amperes per second. I could give more feet per second per second at time = 0, with the empty barrow than with the loaded barrows. The battery of a given voltage could give a bigger ampere per second at time = 0, when inductance (L) was small than when the inductance (L) was larger.

It will be observed then, that the slope, the slant or slope of the previously mentioned three curves at origin is different. The amperes per second (coulombs per second per second) is different. The feet per second per second of the three wheelbarrows is also different at the very beginning of time.

The initial feet per second per second is equal to force divided by mass, F/m . Acceleration = $a = F/m$.

The initial amperes per second when $t = 0$ is equal in analogous fashion to E/L where E is the voltage of the battery and L , the inductance. The resistance in the circuit has nothing to say about the amperes per second(s) at $t = 0$. Only the L controls the $\frac{di}{dt}$ at $t = 0$. The current at $t = 0$ is 0.

At all subsequent instants of time, when the current has attained some value, i , the amperes per second(s) is equal to $\frac{E - Ri}{L}$.

At the very beginning when $t = 0$ and $i = 0$ the acceleration of wheelbarrows (feet per second per second) and the electrical current change (amperes per second) is governed only by mass and inductance respectively. In other words, resistance does not control the time rate of change of current(s) when $t = 0$ since there is no current. Later of course the amperes per second is given by the formula above, $\frac{E - Ri}{L}$. The i is the instantaneous value of the current. The final current is I .

This means that although the current keeps increasing, the amperes per second(s) keeps on decreasing.

The above ideas are graphically portrayed in the 3 curves of Chart 1A, page 203.

DECELERATION

Now finally assume in the human constant push case, that I suddenly decide to stop pushing. Now the kinetic energy stored in the motion of the barrow will again react and try to drag me along. The wheelbarrow is stubborn. It did not want to gain speed and more energy. Likewise, it does not want to lose speed and lose energy. So it rebels, it reacts and tries to keep going, trying to keep up its velocity, its current. I am hanging on so it tries to keep me going and therefore drags me along.

If I decide to suddenly stop entirely, the wheelbarrow does not want to suddenly stop entirely. Its stored energy violently reacts; it provides a big reactive force and throws me down sprawling. There has been a deceleration, a big feet per second per second effect. The entire kinetic energy of the barrow finally ends up in heat of friction due to the barrow and me skidding along the pavement. The stored energy of the barrow has been given back, as it were.

MAGNETIC FIELD AND CURRENT: DE-ENERGIZING

If I decide to open the electric circuit and break the current, the energy stored in the magnetic field of the inductance must do something. The current is decreased and again a reactive voltage is developed. Hence a spark appears at the switch and the stored energy goes down the drain in heat. Some energy is lost in resistance of the connecting wires too.

One final important point is to be made. The inductance reacts with a back voltage only when the current is changing; either increasing or decreasing, when there is a positive or negative coulombs per second per second or amperes per second.

The wheelbarrow reacts with a back force only when it gains speed or loses speed. The amount or the magnitude of the reactive force in such cases depends on the mass or the inertia of an empty barrow or a lead loaded barrow on one hand and the degree or suddenness of speed change, acceleration or deceleration in feet per second per second. The barrow exerted a large reactive force on a sudden stop and there was a lot of deceleration. I was thrown down.

The reactive voltage depends not only on the inductance of the coil but on the time rate of change of increase or decrease of current. On breaking of circuit and sudden stopping of current, the inductance produced enough voltage and quickly gave back its energy to produce a "fat" and very hot spark. This voltage is usually big enough to give one an electrical shock.

NO NAME FOR ACCELERATION IN MECHANICS AND ELECTRICS

In passing it is interesting to observe that in mechanics we do not have a name for a unit of velocity; mechanical current. We have to always say feet per second or miles per hour. Then acceleration too must be expressed in feet per second per second or miles per hour per second.

In electrical affairs we do have a name for a unit current; ampere. So if a current changes we say amperes per second. Of course as has been previously said an ampere itself is a coulomb per second. So changing current is either amperes per second or coulombs per second per second.

To be sure, the "knot" is a unit of velocity in maritime endeavors. A "knot" is one nautical mile per hour or 1.15 statute miles per hour. Therefore acceleration of ships is knots per hour or knots per second.

Maybe someday we will call one foot per second per second acceleration, a Kettering, after Mr. C. F. Kettering of the automobile industry. Still better we might call a mile per hour per second a Kettering; ket for short.

Maybe someday we will call the unit electrical current change or electric charge acceleration, amperes per second(s) (coulombs per second per second), a Carty, after J. J. Carty, a distinguished telephone engineer of yesteryear. In these pages we will denote amperes per second by the letter, "s". The "s" may be zero (0); it may be greater than zero since it changes. Its value at any instant we will call "s". Its maximum value in a particular case will be called, *S*. One ampere per second would be the magnitude of a Carty.

INDUCTIVE REACTANCE

Inductance of the coil is its property to store energy in the form of a magnetic field. That is one fact. The reactive voltage provided by an inductance coil depends not only on the inductance *L*, but also on the time rate of increase of current; the *s*, amperes per second, Carties.

L = inductance, a characteristic depending on the physical dimensions, geometric arrangement of coil and magnetic nature of core material.

$$\begin{aligned}\text{Inductive Voltage} &= L \text{ times time rate of change of current} \\ &= Ls \\ &= \text{henries times amperes per second} = \text{volt} = \text{Henries time Carties}\end{aligned}$$

If a 60-cycle per second alternating current has a maximum value of 10 amperes when $t = \frac{1}{240}$ second, the amperes per second at this time is zero. At initial instant of time when $t = 0$, the amperes are zero but the amperes per second are 3,770. See A. C. graphs on pages 150, 151. The first curve portrays varying values of "i", the second curve portrays varying values of "s". The first curve is a sine curve (a current curve or ampere curve) and the second is a cosine curve, an ampere per second curve, a Carty curve.

Inductive Volage = Inductance times coulombs per second per second.

$$\text{Volts} = Ls = \text{Henries times Carties} = L \frac{di}{dt}$$

$$\begin{aligned}\text{Mechanical Force} &= \text{Mass times feet per second per second} = M \frac{dv}{dt} \\ &= \text{Mass} \cdot a = \text{Mass times acceleration} = \text{Pounds times Kets}\end{aligned}$$

L, *s*, and *Ls* are different factors; they involve specifically different concepts. The *L* of a coil is independent of current or current changes. It is a source of back voltage only when there is an "s".

The wheelbarrows were loaded differently and each had a mass in its own right whatever the acceleration. Each exhibited a reactive force only when it was accelerated or decelerated.

Mass of a physical object is validly analogous to inductance; mass is not a cause and mechanical reactance or force a result. Mechanical reactance or force needs both mass and acceleration.

THE HENRY DEFINED

So now we are ready to answer the reader's question, namely: How can we find the amount of inductance an inductor has; how can we know what a henry is?

Assume we know the time rate of change of current at some particular instant to be one ampere per second; one coulomb per second per second. The circuit might be a d-c or a-c circuit. If the back or reactive voltage at that same particular instant is 1 volt, then the coil would have 1 henry of inductance. In other words, one henry is the inductance that gives rise to a back voltage of one volt when the current is changing at the rate of 1 ampere per second, one Carty.

If, at the instant of opening of a switch in some other case, the decrease or time rate of change of current " s " was 50 amperes per second and the voltage at that same instant was 250 volts, then the inductance would be 5 henries. The " s " might some time be 1000 amperes per second. If " L " were 5 henries, there would be a reactive voltage of 5000 volts. It might be short-lived, but it would be a very real voltage.

If the source of energy d-c or a-c had been connected to a circuit with only resistance in it and no inductance, then the current in amperes would instantly become the value decreed by Ohm's Law. If a battery of 12 volts were connected to a circuit of 12 ohms, then the current would be almost immediately 1 ampere. This one ampere would be the current, a millionth of a second or so after the closing of the switch. The energy supplied (RI^2t) to the resistance would immediately be lost in heat, and so it would be all the time the circuit was closed. At the instant of opening the switch, there would be no current and no spark at switch since no energy had been stored in the system. Of course to have a circuit without a wee bit of inductance is impossible.

With an inductance in a d-c circuit and an a-c, too, the reactive voltage resulting from the fact that energy is being stored tends to keep the current down for a little while, Chart 1, Page 181. Assume, in the above example, one had an inductance 10 henries in the circuit with the 12 ohms. Assume, also, that the wires of the inductance coil itself had absolutely no resistance. This could not be but play nevertheless, or assume it was the case.

When the switch was closed, the reactive voltage (Es) of the 10 henries (L) of inductance and the " s " kept the current from rising instantly. After some small time, the current " i " might be a .01 ampere; later, after some small additional time, it might be .1 ampere; later, .5 ampere; later, .8 ampere; and finally, 1 ampere.

The corresponding "back" voltages at the assumed instants of time, $\left(\text{the } L \frac{di's}{dt}\right)$ would be equal to $E - ri$; namely 11.8 volts, 10.8, 6.0, 2.4 and 0 volts. The $\frac{di's}{dt}$ would be equal to $\frac{E - ri}{L}$; namely 1.118 amperes per second, 1.08, .60, .24 and 0 amperes per second.

At the instant the switch was closed and the current was zero, it was changing at the rate of 1.2 amperes per second; $s = \frac{Et}{L}$. At the instant it was .1 ampere, its time rate of change would be something less. When it was .5 amperes, its time rate of increase would be still less. When the current was .8 ampere, its time rate of change would be still less. The " s 's" were different as were the i 's. The "back" voltages also changed; they gradually decreased.

All the time, the current was increasing but its time rate of change was decreasing; the " s " was decreasing.

$$\frac{di}{dt} = s = \frac{Et}{L} = \frac{E - ri}{L}$$

Finally when the current had risen to 1 ampere, there were no amperes per second change, and hence no reactive voltage.

This final 1 ampere would be the same as in previous case of no inductance. It took 8.3 seconds ($10 L/R$) for the current to grow, as it were, to the final one ampere when there was inductance. If inductance were larger, say 20 henries, the final current would be the same but 16.6 seconds would have to elapse before the steady state was arrived at.

During these assumed elapsed times, the reactive voltage of the inductance and the " S " was opposing and holding down the increase of current. But at the end, there would be 1 ampere through inductance and resist-

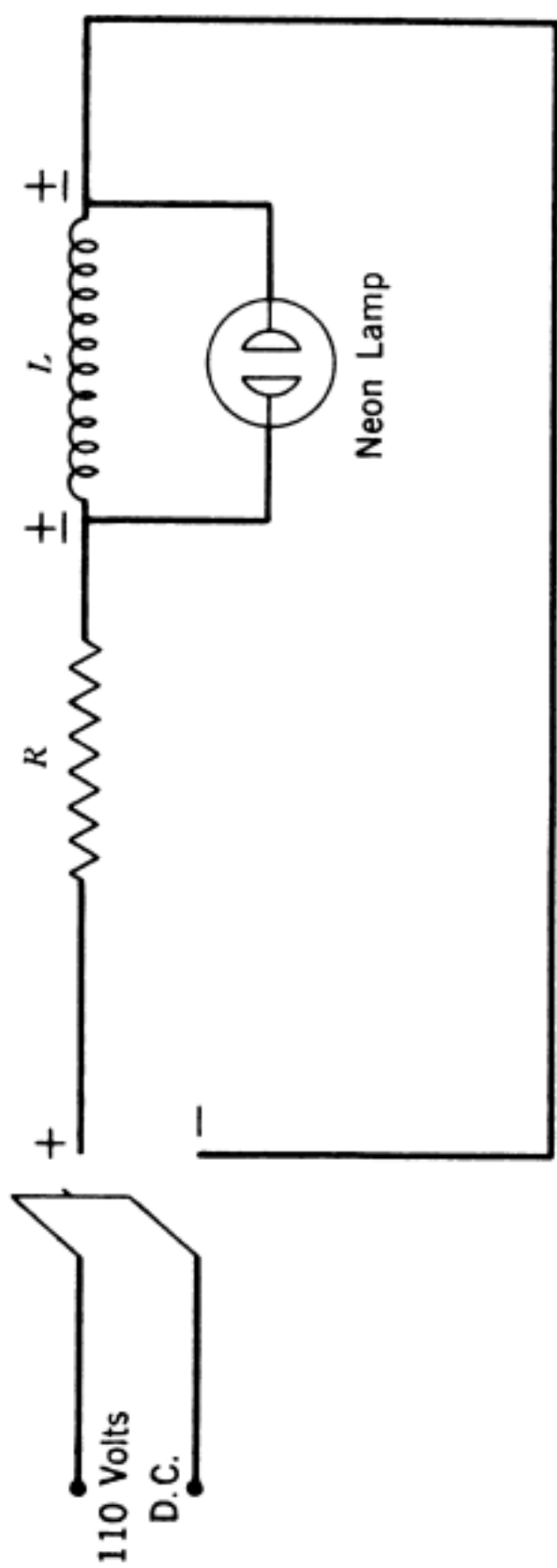


FIGURE 55

ance. All during the 8.3 seconds, required to rise gradually from zero to 1 ampere, energy was being stored up in the magnetic field around the inductance coil. It will be seen in the Charts 17, 17A, page 223, how the current in the circuit increases. Chart 1, page 181. The increase is not the same in the several cases portrayed. See Charts 1A, 1B, and 1C, pages 203, 204, 206.

The increase in current (the amperes per second, " s ") is large at the beginning but the increase slows down after a while and at the end of 2 seconds practically stops increasing. At beginning the " s " is E/L as previously discussed. The current is always increasing during the transient period but it increases at a decreasing rate. It would be well for the reader to read this last statement several times. The statement that the current increases at a decreasing rate tersely expresses an important truth. Its corollary is, the back or reactive voltage (Ls) decreases at a decreasing rate. The changing slope of the " i " curve, Chart 1 (page 181) and of the e_L curve Chart 2 (page 185) indicate the above.

Inspection of the current and reactive voltages curves in the Chart 2, page 185 and Chart 4, page 190 will further reveal the validity of the above statements.

The curves show very nicely that the reactive voltage (Ls) is maximum (LS) when the current is zero but changing and that the reactive voltage (Ls) is zero when the current is a maximum but is experiencing no change. The " s " is zero when the steady state is arrived at.

AN INTERESTING EXPERIMENT WITH INDUCTANCE

The accompanying sketch Figure 55 shows a simple circuit arrangement of resistance and inductance in series connected to D. C. voltage source. A two electrode neon flash lamp is connected across the inductance. This neon lamp is a sort of quick acting D. C. voltmeter. It also shows instantaneous polarity or direction of voltage. When connected to a constant 110 volt D. C., one or the other terminal will constantly glow, depending on connection. The terminal or half disc having positive voltage will glow. If connected to 110 volt A. C., both terminals will flash but alternately. This alternating glow can be seen readily if the neon lamp is looked at in a tilting mirror.

When the switch is closed with the D. C. supply furnishing the energy and the increasing current, the left hand side terminal flashes momentarily. This flash reveals not only the existence of a back voltage but also the polarity of the back voltage. It is opposite to the voltage applied to the inductance. After the current has stopped increasing, the back voltage is zero and the neon lamp goes out, since the RI drop across the inductance is not large enough to ignite the lamp.

The reactive voltage has tried to keep current from rising and going clockwise in the circuit. So as said before, the left hand side of the inductance is momentarily positive and the left hand terminal of the neon lamp flashes momentarily.

When the circuit is opened and the current tries to decrease, the neon lamp reveals the brief existence of a reactive voltage which tries to keep the current going the same direction, namely clockwise. See sketch. The right hand terminal of the inductance is now positive and the right hand terminal of the neon lamp flashes momentarily.

The reactive voltage due to the inductance and the " s " is brief in each case and is opposite in direction on make and break. The neon lamp has played its clever role as a voltmeter and a polarity indicator.

ALTERNATING CURRENT, INDUCTIVE REACTANCE

The value of an alternating current is continuously changing in magnitude. Its direction also changes. See accompanying graph (Figure 56) of an alternating current. At one instant the current is zero; then it rises to a maximum; then decreases to 0. It follows the same pattern during the negative half of the cycle: increases to a negative maximum and finally comes back to zero.

The accompanying sketch shows an alternating current whose frequency is assumed to be 60 cycles per second. Its period (T) or time of one cycle is therefore $\frac{1}{60}$ of a second.

It is assumed in the example portrayed, that the current rises from zero to a maximum and instantaneous value of 10 amperes. This would happen in $\frac{1}{4}$ of a $\frac{1}{60}$ of a second namely $\frac{1}{240}$ second. In a simple A. C. gen-

erator this would result from a 90° ($\frac{\pi}{2}$) rotation of the armature. Inspection of the curve shows that in $\frac{1}{120}$ ($\frac{3}{240}$) of a second, the current is again 0, corresponding to an armature rotation of 180° (π). Then the current increases to a negative maximum of 10 amperes after $\frac{1}{80}$ ($\frac{3}{240}$) of a second and a rotation of 270° ($\frac{3\pi}{2}$). Finally after one complete cycle of $\frac{1}{60}$ second, (360° rotation, 2π) the current is again 0.

The 60 cycle A. C. curve in this sketch merits a lot of inspection and study. The instantaneous values of current (i) at different values of time (t) are shown and the instantaneous "s's" at corresponding times (t) are shown, namely the amperes per second.

It is of particular interest to realize that in every case of an alternating current whose maximum value of 10 amperes, 1 ampere, or 50 amperes (the Carties), the amperes per second(s) when the current is zero, is $2\pi f$ times 10 ampere per second (3,768), $2\pi f$ times 1 ampere per second (376.8), $2\pi f$ times 50 amperes per second (18,840) respectively.

A second sketch (Figure 57) accompanies this part of the discussion. It is the amperes per second(s) curve. The current curve is a sine curve: the time rate of change of the sine current curve is amperes per second, a cosine curve.

As the current itself varies between 0 and the maximum I , there will be different Carties, amperes per second(s), at each different value of current, i . Take a look at both attached curves.

It will be seen from first curve (Fig. 56) that during the first quarter cycle, the current increased from 0 to 10 amperes. The rate of increase (the amperes per second), the (s) decreased from $2\pi f$ times 10 amperes per second (3,756) to zero amperes per second.

One might very appropriately ask at this point how does the factor 2π or 2π (6.28) get into the act? The answer is that when a simple A. C. generator produced an alternating current, the coil of the single loop of armature has been rotated at a certain angular velocity (ω) (omega) and turns one complete circle; 360° of angle or 2π radians of angle per second.

i, s , SINE, COSINE

The first curve of the two accompanying curves is the A. C. current (i) curve. It graphically portrays how the current itself in amperes rises and falls as time progresses. The curve is a sine curve. Its equation is $i = I \sin \theta = 10 \sin \theta = 10 \sin \omega t$.

The second curve (Fig. 57) is not the current (ampere) curve but is the ampere per second curve. It sketches out the time rate of change of current(s) at each given instant of time. It graphically portrays the varying values of (s) at different instants of time. It is suggested that this second curve be looked at carefully in comparison with the first curve. The equation of this curve is $s = S \cos \theta = 10.2\pi 60 \cos \theta = S \cos \omega t = S \cos 2\pi f t$.

It can be said appropriately that the " s " curve is 90° different in phase compared to the i curve. The s curve (the amperes per second), can be said appropriately to lead or to be in advance of the " i " curve. It is already a maximum (S) at the initial instant when the current (i) is zero.

The " s " curve is a cosine curve as compared to the " i " sine curve. At $t = 0$, s is already a maximum, S . At this time $t = 0$, $i = 0$. Hence, the " s " curve leads by 90° or $\frac{\pi}{2}$ radians the " i " curve. The reactive voltage Ls (in general), is a maximum at $t = 0$ and leads the current i by $\frac{\pi}{2}$ radians or 90° .

RADIANS

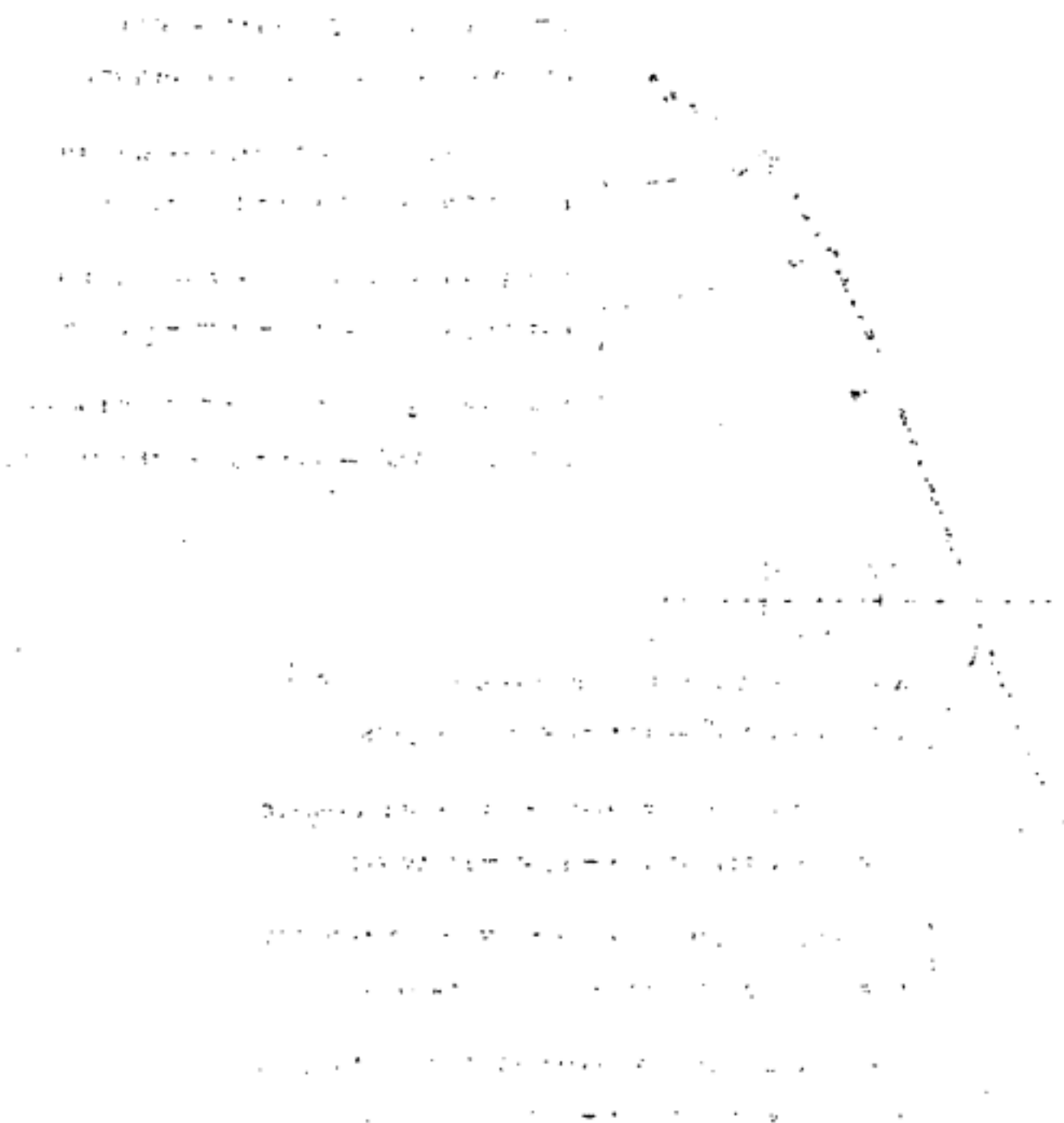
At every instant of time, the value of the current (as is seen in 1st curve (Fig. 56) depends on the angle of rotation in general and on the trigonometric sine of the angle in particular. Simple trigonometry is a good mathematic tool to tell what the amount of current is. Trigonometry and Faraday's principle of production of voltage by cutting of the lines of magnetic flux, provided by the permanent magnets of the generator, decree that the angle cannot be expressed in degrees (30° or 90° etc.) but must be expressed in terms of angle measured in radians. One radian of angle = 57.3° . In a complete circle there are 360° but 2π radians the 360 is purely

i =
=
s =
=

i =
=
s =
=

i =
=
s =
=

i

$\Delta =$ $=$ $=$ $i =$ $\Delta =$ $i =$ $\Delta =$ $i =$ $\Delta =$ $i =$ 

arbitrary and conventional number assumed by man, but the 2π radians is a simple basic relation in the geometry of a circle. $180^\circ = \pi$ radians and $90^\circ = \frac{\pi}{2}$ radians. Therefore simple but not arbitrary principles decree that 2π gets in the act when we wish to know the changing instantaneous amounts of voltage or current during the complete cycles.

The trigonometric circular sine of zero (0) angle is zero and is unity (1) when angle is 90° or $\frac{\pi}{2}$ radians. The circular cosine of zero (0) angle is unity (1) and is zero (0) when the angle is 90° or $\frac{\pi}{2}$ radians.

HOW TO FIND "s"

As stated above, the value of the current itself depends on the angle of rotation in general and on the trigonometric sine of the angle in particular.

$$i = I \sin \theta; \quad \theta = \text{theta} = \text{angle}$$

At this point a good friend and ally, differential calculus lends its talent and decrees that the amperes per second depends not on the sine of the angle but the cosine of the angle. Calculus says that the derivative of the sine is the cosine. $\frac{d}{dt} \sin 2\pi ft = 2\pi f \cos 2\pi ft$. The slope (s) of a sine curve at every instant provides a point on a cosine curve.

This fact provided by the calculus means that the time rate of change of numerical value of the sine at each instant of time is numerically equal to a constant times the cosine at that instant. The two previous curves graphically portray and illustrate this simple and basic truth.

The sine of zero angle is zero but the cosine of zero angle is unity (1). The sine of 90° or $\frac{\pi}{2}$ angle is unity (1) but the cosine of 90° or $\frac{\pi}{2}$ is zero. These are simple facts and relations supplied by circular trigonometry.

Putting the above words in symbols.

$$\text{Instantaneous Current} = \text{Amperes} = i = I \sin \theta = I \sin \omega t = I \sin 2\pi ft$$

where I = maximum current

$$\text{Amperes per Second} = s = 2\pi f I \cos \theta = S \cos \theta$$

On the chart, the s is shown to be at $t = \frac{1}{480}$ second as 2π 60 times 8.6 amperes per second. This is the case because the angle is 30° or $\frac{\pi}{6}$ and the cosine of this angle is .86. The current itself at this instant is 5 amperes. This is true because sine of the same angle is .5.

At $t = \frac{1}{480}$ second, the current itself is 7 amperes and the "s" is 2π 60 times 7 amperes per second (2,537). It happens that at $t = \frac{1}{480}$ second the number 7 appears in both places. It happens because $\sin 45^\circ$ and $\cos 45^\circ$ are equal, namely .7.

If an alternating source of a given voltage is connected to a circuit consisting of a resistance and an inductance, the energy is delivered to the circuit. The energy supplied to the resistance is irrevocably lost but the energy supplied to the inductance is stored during certain parts of the cycle and returned during other parts of the cycle. The amount of stored energy is, of course, equal to the energy that can be gotten back.

The current flowing in the circuit is varying (increasing, decreasing and changing direction). At each and every instant of time in its variation its value is determined by the constant resistance and instantaneous value of back voltage due to the inductance.

The instantaneous back voltage has different values at each instant because the current is changing at a different time rate at the different instants of time; not only are the amperes different but also the amperes per second are different; the Carties are different. The amperes per second(s) will always be $2\pi f I \cos \theta$. These values of "s" have been designated in the second curve. (Figure 57).

So in A. C. one meets different currents and different time rates of change of current at every instant during the cycle. If one has 60 cycle A. C., then the whole array of current changes take place in $\frac{1}{60}$ of a second. If one had 120 cycles, then the whole array of change would take place in $\frac{1}{120}$ of a second. Obviously, if the overall change happens in $\frac{1}{120}$ of a second, then each individual instantaneous change (amperes per seconds) would be made in a correspondingly shorter time. If one had 500 cycles, the whole array of changes in current would all happen in $\frac{1}{500}$ of a second, then each individual instantaneous change would happen in a correspondingly shorter time.

If the current through the inductance L changes at a fast rate and a slower rate, the instantaneous back or reactive voltages (Ls) would correspondingly vary; they would be great and small, very great and very small, depending on the cycles per second. As stated before, the 2π and the cosine factors always enter the picture as does f , the frequency.

As per above argument about 60 cycles and 120 cycles, the rates of change of current at 120 cycles would be greater on the average than at 60 cycles. The reactive voltages in turn with a given L would also be greater on the average when A. C. is 120 cycles than when A. C. is 60 cycles because the overall " s " at 60 cycles per second would not be as great as at 120 cycles per second.

In any case of whatever frequency, the reactive voltage of L is greatest when current is 0, because at the instant when A. C. current is 0, the current is actually changing at its fastest rate, the " s " is greatest (S) since $\cos \theta$ is largest, namely unity. In any case of whatever frequency, the reactive voltage is zero when the A. C. current is a maximum (positive or negative) because at that instant, $\left(\frac{\pi}{2}, 90^\circ\right), \left(\frac{3\pi}{2}, 270^\circ\right)$ the current is not changing at all; the cosine of the angle is zero and the " s " is zero. See again the curve of alternating current in " i " and " s " curves in accompanying sketches.

The above situation prompts the frequently heard and completely valid and readily understood statement, that, in a circuit with an inductance, the current in the inductor lags 90° behind the back or counter or reactive voltage developed in the inductor. This idea was encountered in the experiment mentioned on page 148, Figure 55.

The gist and the "wrap up" of the above several paragraphs is that the overall average reactive voltage (Ls) will be greater when the A. C. is 120 cycles per second as compared to 60 A. C. and greater at 400 cycles per second as compared to 120 cycles per second and of course in turn to 60 cycles per second. The inductance L itself is a fixed factor independent of the frequency. The inductive reactance (Ls), depends on frequency which in turn controls time rate of change of current, " s ".

Since A. C. means alternating or changing, the above term overall average must be used in comparing A. C. to D. C.

In direct current systems $E = RI = \text{Voltage} = \text{Resistance times Current}$.

In an A. C. system, e also $= Ri$ but i is now the instantaneous value of the changing alternating current.

In alternating current on the average; on the overall is $E_L = \text{Reactive Voltage} = \text{Inductive Reactance times Overall Current, } I_e$.

Inductive Reactance is often given the notation $2\pi fL$, or X_L .

$$\text{Inductive Reactance} = X_L = 2\pi fL = L\omega$$

The $2\pi f$ is sometimes called omega (ω) which is really the angular velocity measured in radians per second; $\omega = 2\pi f$.

In sinusoidal A. C., the overall or effective value current during the entire cycle is .71 of maximum current, $I_e = .71I$. The word effective is appropriate because it is the value which in an alternating current circuit produces the same result as a direct current of the same value when one considers the heat relations in formula $I_e^2 Rt$.

The formula $E_L = \omega LI_e = 2\pi fLI_e$ states in symbols the truth that was argued about in words above; namely that the overall inductive reactance voltage depends not only on the inductance L , but also on the frequency at which the current changes, the amperes per second, and finally on the effective value of the current, I_e . Increased frequency means that changes take place more rapidly. It is hardly tenable or legitimate

to say that L is the cause and reactive voltage, E_L the effect. The cause involves both, L and f . The f decreases the "s" (amperes per second). Still better, the cause relates to the fact that energy is being put into a system and stored up in a magnetic field. A measure of the amount of energy and the inductive reactance is not only dependent on L but also the time rate at which energy can be put in which involves f , the frequency. The cause is really a basic electromagnetic phenomenon, in its overall qualitative aspect, namely the storage of energy. The quantitative aspect or reactive voltage involves f , the frequency, L , the inductance and the effective current, I_e . Again these combined factors of quantitative relations may be put in symbols.

$$E_L = 2\pi f L I_e = L\omega I_e \text{ (where } \omega = 2\pi f \text{ radians per second)} = X_L I_e.$$

Volts = Inductive Reactance, $L\omega$ (Ohms) time Effective Current, I_e (Amperes).

Again overall or equivalent effective value of current, $I_e = .71$ maximum current. This effective value or overall value of current (I_e) being .71 of maximum current comes from a simple truth and arguments that can be obtained from some text book. Of course, integral calculus really supplies the answer.

The doctrine; the basis and the approach for finding the effective value of alternating current are that it will be equivalent to that direct and continuous current which in a given time (1 cycle for example) would provide the same amount of heat as would all the various and many instantaneous values of alternating current in the same time. Since the heat depends on the square of current, the effective value or equivalent direct current value is the square root of the sum of the squares of all the varying alternating currents. As said before, the calculus is the beautiful tool which provides an explicit answer.

CAPACITORS AND CAPACITIVE REACTANCE

Inductance was defined as that property or quality of physical things which permits of translation of electrical energy into the energy of a magnetic field. It was pointed out that inductance was nicely and validly analogous to the mechanical mass of an object.

By the same basic sort of approach, capacitance may be defined as that property or quality of a physical system which permits of translation of the electrical energy of a battery or generator into the energy of electric field. The physical system in this case is any physical object on which a charge of electricity can be placed by a supply electrical voltage and energy source.

The pair of wires of a telephone line, a trolley wire and track, the two electrodes of a storage battery, etc., etc., all comprise a capacitor which can be electrically charged. Usually we think of a capacitor as two metal plates or sets of parallel plates separated by a dielectric of air, glass, paraffine, mica, etc., but every physical object which can be charged has capacitance.

ELECTRICAL CAPACITORS AND COILED SPRINGS

It is to be pointed out at once that a capacitance can be nicely and validly thought of as analogous to a coiled mechanical spring.

An electrical capacitor stores energy in terms of electric field. A capacitor is a reservoir of electric field energy. This energy has the quality or nature of potential energy. This capacitor energy can be kept stored for quite a while and then can be recovered or gotten back.

A compressed or stretched spring stores energy, a sort of potential energy. This energy can be stored in the spring for quite a while and can be later recovered or gotten back.

When an electrical capacitor is energized or charged by connection to a battery, the electric current (amperes) into the capacitor is a maximum at beginning of charging time and is zero at end. The "back" voltage acquired by the capacitor is zero at beginning ($t = 0$) when current is a maximum. The "back" voltage acquired by a capacitor is maximum at the end of charging time (T) when current is zero. That the current into capacitor leads by 90° the back voltage acquired by the capacitor is an often repeated and readily understood truth. The current is already a maximum (I) when $t = 0$ but voltage on capacitor is 0; hence, as stated, the current leads the "back" voltage acquired by the capacitor.

When a supported spring, a mechanical capacitor, is energized (stretched or compressed) by some outside

force, namely a human being, the "mechanical current" or the velocity of the movable end of the spring is a maximum at the very start when the outside pull or push is applied. At the very instant of start, the back pull or push "aroused" in the spring is zero. As the outside force (the mechanical battery) keeps working, the back pull or back push of the spring increases until the spring is fully "charged" or energized. The back push of the spring is equal to applied push. At the end, the mechanical current (velocity of movable end of spring) is zero and the back pull (voltage) aroused in the spring is a maximum and equal to applied pull or push. Here again the "current" or displacement in the spring leads the applied force by 90°.

Mechanical velocity is validly analogous to electric current. Electric current is really charge times velocity; it is QV ; charge per unit time; coulombs per second.

The analogy between an electrical capacitor and a spring in the above particular respects is completely pertinent and valid.

The capacitance of an electrical capacitor (farad) is the ratio of its electric charge (coulombs) to the voltage (volts) acquired by it due to its charge, Q .

$$C \text{ (Farads)} = \frac{Q \text{ (coulombs)}}{V \text{ (volts)}}$$

The Q in coulombs is the amperes times seconds ($Q = It$) and is the strain or yield, as it were, which has been put into the capacitor. The applied volts are not only the stress to which the capacitor has complied or yielded but also the stress "aroused" in the capacitor itself.

So one could very validly say:

Capacitance is ratio between electrical strain and electrical stress. If, at one volt, a capacitor had 100 coulombs of strain or charge, it would be a big capacitor and a compliant and receptive capacitor in which a lot of energy was stored with only one volt of stress involved. The capacitance of such a capacitor would be fantastically large: 1 farad. The stored energy would be $\frac{1}{2} CV^2$ or $\frac{1}{2} QV$ or 50 joules (50 watt seconds).

This big capacitor, this very receptive and compliant capacitor, would have a large ratio between strain and stress. It would be said to have a large coefficient of compliance (C), usually called capacitance, and is expressed in farads or microfarads.

MECHANICAL CAPACITANCE

The "capacitance" or the compliance of a coiled spring is the ratio of elongation, strain, or mechanical "charge" to the back force acquired by it when stretched by an outside force. It is the coefficient of compliance, the coefficient of "yieldability".

The mechanical "charge" is the velocity (current) times the seconds involved in the period of charging (pulling). The "charge" is the strain (the elongation) in the spring produced by one unit of mechanical force applied. In mechanics, velocity multiplied by time equals a distance. The "charge" or strain of a coiled spring is the number of inches that the spring stretches when one unit of force (say, 1 pound) is hung on the spring. If a spring stretches a lot for one pound of force pulling it, then it would be a big mechanical "capacitor". It would be a very compliant spring. A lot of energy would be stored in it by only one pound of force. The big electrical capacitor with a large coefficient of compliance (capacitance) stores a lot of energy with only one volt involved.

The spring which stretches a lot of inches and provides a lot of strain per unit of force would not be a stiff spring but would be a compliant spring.

The ratio of strain to stress, the ratio of elongation to pull, can appropriately be called the coefficient of compliance (K) of a spring; the coefficient of "yieldability." The " K " has no simple name, but is the inches per pound. A " K " of one inch per pound might be called a "mechrad" (a mechanical farad) or maybe a Fulton after Robert Fulton, of steamboat fame.

The coefficient of compliance of an electrical capacitor is generally called its capacitance (C) = $\frac{\text{strain}}{\text{stress}} = \frac{Q}{V}$ equals Coulombs per Volt equals Farads.

The coefficient of compliance of a "mechanical" capacitor has no particular name but only a symbol $K = \frac{\text{strain}}{\text{stress}} = \text{Inches per pound}$.

Electrical strain and stress in one case and mechanical strain and stress in the other are validly analogous.

The analogy between an electrical capacitor and a coiled spring (a mechanical capacitor) in the above respect is completely pertinent and valid.

CHARGING A CAPACITOR

Assume one connects a source of energy with constant voltage E to the metal plates of a capacitor. In the circuit a resistance is also connected. When the circuit is closed, the final situation is that the plates of the capacitor have a voltage of E between them. The more important fact is, however, that some energy supplied by the source has been delivered to the capacitor. The capacitor is a reservoir. Again it is to be emphasized that the energy delivered is stored as an electric field in the dielectric between the plates ($\frac{1}{2}CV^2$) and is recoverable. The energy delivered to the resistance in the charging process is lost in heat.

It takes time, measured in seconds or fractions thereof to achieve the storage of energy in the capacitor. The charged capacitor is a reservoir of energy.

It is in order therefore to inquire what happens qualitatively and quantitatively during the period of entire time of energy storage; during the transient period of time. In discussing the quantitative aspects, one must again ask questions about voltage, current, pure ohmic resistance, capacitance, amount of energy stored and capacitive reactance. An inductive reactance was talked about in ohms, so capacitive reactance will be talked about in ohms. But inductive reactance ohms and capacitive reactance ohms do not involve loss of energy in heat, but involve storage of energy in watt seconds, watt hours or milliwatt milliseconds. Energies stored in some capacitors might well be so small that their storage energy might well be talked about in microwatt microseconds. This latter amount of energy would be so little that it would be difficult for one to appreciate. It might be even smaller than that expended by a mosquito flying .1 inch. However in electrical communication, this small bit of stored energy might play a very important role in some operating television, radio or wire telephone apparatus.

RESISTANCE OF A CAPACITOR

The term resistance of a capacitor is really not a resistance in the ordinary sense of the term but the capacitor does perform at times in a manner suggestive of a varying resistance. At the instant the switch, in the above mentioned circuit of voltage, (E , R and capacitor C), is closed, the capacitor acts as a zero resistance. The current flowing in the circuit at that initial instant and continuing for a microsecond or a micromicrosecond is a maximum and equal to E/R . In other words, at the beginning instant of time, the current is at once a maximum E/R , just as if there were no capacitor. The initial surge or rise of current depends entirely on the resistance in the circuit; the capacitor has nothing to say about it. See accompanying circuit sketch Figure 58 and Chart 5, pages 192, 207, 258.

ROLE OF INDUCTANCE AND CAPACITANCE AT ZERO TIME

In passing, a very interesting and important idea relative to the difference in initial happenings when an inductance and capacitance are energized deserves mention.

On the instant of connecting a battery to an inductance, the resistance in the circuit has nothing to say about the happenings at that first instant. The inductance is in the driver's seat and completely controls the time rate at which the current begins to increase; the amperes per second(s). Of course resistance enters the picture later and finally has everything to say about the current.

At the instant of connecting a battery to a capacitor, the resistance in the circuit has everything to say about the current at that instant. The capacitor has nothing to say about happenings at this very first instant. Of course later, the capacitance enters the picture and decrees that the current becomes zero.

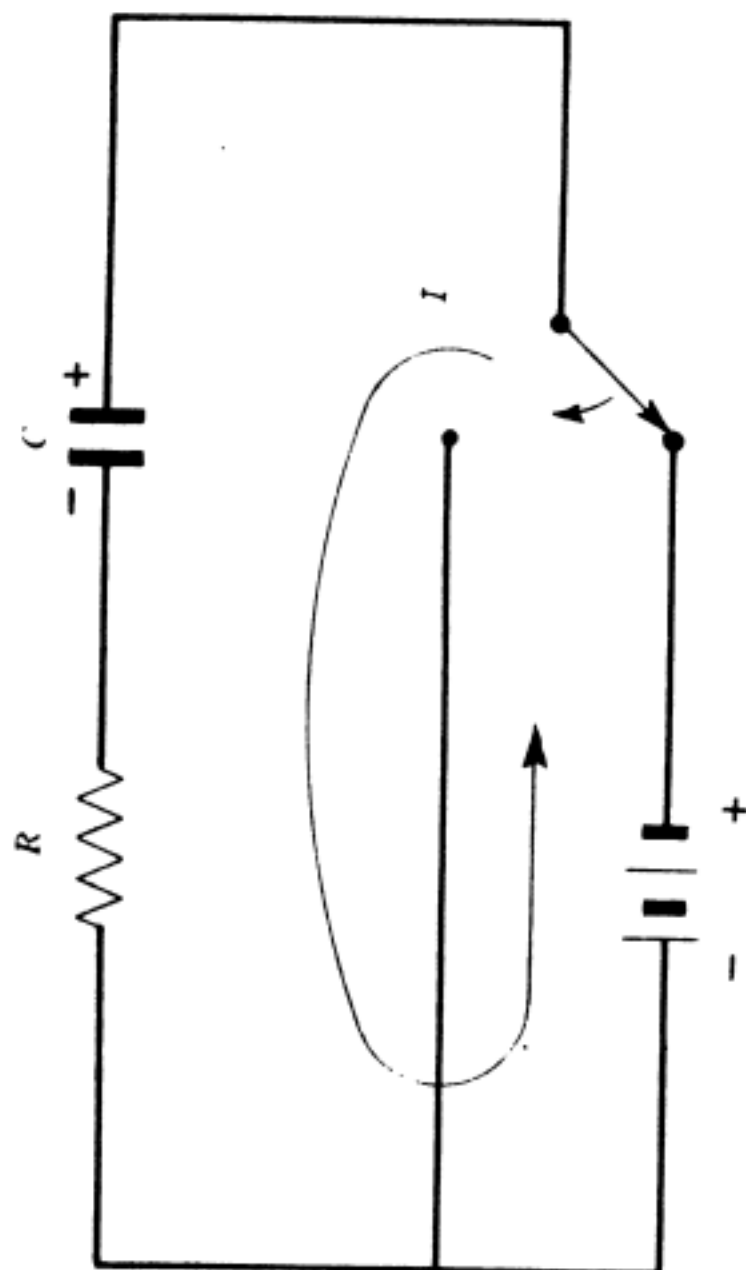


FIGURE 58

HAPPENINGS ON CHARGING A CAPACITOR

During the very first microsecond, a bit of charge has been acquired by the capacitor. This bit of charge on the capacitor plates means that the plates have acquired a voltage which is a back or counter or reactive voltage. The two voltages, that of source E and capacitor plus the Ri drop oppose one another at all times as the current decreases. The same argument applies at each instant of ensuing time, so when the voltage acquired by capacitor is finally equal to E the applied voltage, the current in the circuit is 0. At this final state of affairs the electric charge or quantity of electricity, Q , on the capacitor is CV . There no longer is any Ri drop because the current has decreased to zero.

The Charts 6, 7 pages 194, 195 graphically and convincingly illustrate the train of events.

In one micromicrosecond or so after switch is closed, the current rises from 0 to its maximum E/R , according to Ohm's Law. The voltage (e) and charge ($q = It$) on plates of capacitor at the very beginning is 0. In the ensuing microseconds or milliseconds, the current gradually decreases to 0 but the voltage and charge (q) acquired by capacitor gradually increases to E and Q . During the first few milliseconds, the time rate of decrease of current (amperes per second) is very rapid. The current continues to decrease but does so at a decreasing time rate. The " s " decreases. The acquisition of voltage by the capacitor increases at a rapid time rate at the beginning but its voltage acquisition increases at a decreasing rate. The reader ought to read these last three sentences several times and in between readings look at the curves. Furthermore the reader would do well to go back to the pages where a similar sort of statements were made about matters in an inductance circuit.

After the transient state is over and the steady state is reached, the result is that energy has been stored as an electric field in the dielectric of the capacitor. The amount of energy stored is given by formula, Energy = $\frac{1}{2} CV^2$.

$$\text{Energy (Watt Seconds)} = \frac{1}{2} C \text{ (farads)} V^2 \text{ (volts)}^2$$

Also after the shooting is over, the capacitor has a charge Q (Coulombs) and has acquired a voltage V equal to the voltage of the charging battery. $Q = CV$.

$$\text{Also therefore, Energy (Watt Seconds)} = \frac{1}{2} Q \text{ (Coulombs)} V \text{ (Volts)}$$

It is to be noted that one coulomb of electrical charge = 6.28×10^{18} electrons. An ampere is 6.28×10^{18} electrons per second.

A coulomb is one ampere second. An ampere is a coulomb per second.

CHARGING CURRENT LEADS VOLTAGE ACQUIRED BY CAPACITOR

It is to be pointed out that, at this point of the discussion of capacitance, one can ask more good questions which are nicely answered by simple arguments.

In the inductance case, the back voltage of the inductance was a maximum when the current was zero. This initial back voltage is entirely dependent on L ; R plays no part. It was stated therefore that the current lagged the back voltage "aroused" in the inductance.

In this present case of a capacitor circuit, it is readily appreciated that the back voltage acquired by capacitor lags the current. In other words, the charging current into the capacitor leads the back voltage acquired by the capacitor. The current in the capacitor is a maximum at beginning when the capacitor has acquired neither a charge nor a back voltage. There is no mystery, no high brow stuff and there need be no misgivings, no lack of understanding about the two cases. Careful thinking must be applied to be sure but all is "according to Hoyle"; all is according to simple facts and truths; all is according to a basic law; it all makes sense.

DEFINITION OF FARAD

The previous arguments, ideas and principles provide a good basis for a definition of a farad, the unit used to measure capacitance of a capacitor.

It is to be remembered that at times during the charging period, the current (i) is changing (decreasing), but the charge (q) and the voltage (e) acquired by capacitor is also changing (increasing).

If the current into the capacitor at some instant is one ampere, the capacitor would be acquiring a charge at the time rate of one coulomb per second at that same instant. If the voltage acquired by the capacitor is

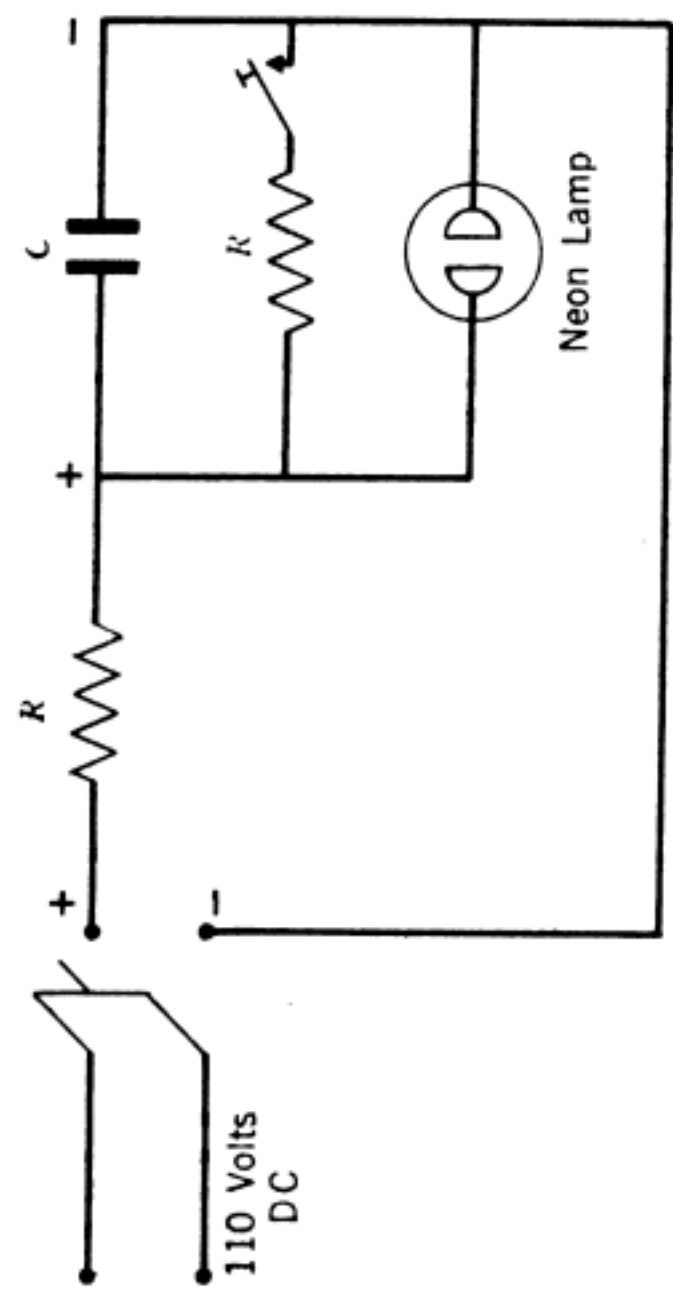


FIGURE 59

increasing at the rate of one volt per second at that very same instant, then the capacitance under such conditions would be one farad. This definition of a farad might be called a dynamic definition. A "static" definition states that if the charge on a capacitor is one coulomb and its potential is one volt, the capacitance is one farad. The capacitance of a capacitor is not the charge (coulombs) it can hold but is the charge (coulombs) per volt that the capacitor makes possible. The maximum charge and voltage it can hold depends on breakdown voltage nature of the dielectric.

One can therefore say that the current (amperes or coulombs per second) flow into a capacitor depends upon the applied volts and the farads and foretells the time rate of rising potential (volts per second) acquired by the capacitor.

AN INTERESTING EXPERIMENT WITH A CAPACITOR

If a circuit arrangement is made with a capacitor substituting for an inductance as in a previous case, some points argued about above are clearly illustrated. See accompanying circuit sketch, Figure 59. At the instant the switch is closed, the neon lamp will not glow at once (the elapsed time is very short and may not be observable) but only after a little while when the capacitor has acquired sufficient charge and voltage. It will however, in contrast to inductance case, stay lighted as long as switch is closed and even quite a while after battery is disconnected. Since events are happening very quickly, the neon lamp will appear to light up instantly but it really does not. When the main switch is opened, the left terminal of the capacitor will stay positive and the neon lamp disc connected thereto will glow. This means that the left terminal of the capacitor is positive both times. The lamp will glow for a little while until all the capacitor energy has been converted to light energy. If the capacitor short circuit switch is closed, its stored energy will be converted to heat and the neon lamp quickly goes out.

In the inductance case, the reactive voltage on opening of circuit tries to keep the current going in the original direction. In the capacitor case, the current on discharge comes out in opposite direction to that which it went in, so left terminal is positive on both closing and opening of circuit and same neon disc glows accordingly. To put this fact in another way, the current on closing and opening of switch in an inductive circuit flows in same direction. In a capacitance circuit the current on opening of circuit flows in a direction reverse to the charging current. See circuit sketch, Figure 59.

A. C. AND CAPACITIVE REACTANCE

The inductance reactance of an inductance in an A. C. circuit was shown to be $X_L = 2\pi fL = \omega L$.

This means that a larger L and a higher f (frequency) makes for greater inductive reactance to restrict flow of A. C. current. Ohm's Law in an abridged form becomes $I = \frac{E}{\sqrt{R^2 + X_L^2}}$.

If one considers the capacitive reactance, one readily comes up with the comparable sort of formula $X_C = \frac{1}{2\pi fC} = \frac{1}{C\omega}$.

This means that if C is bigger and f is higher, the overall or effective value of current in the capacitor will be bigger. The situation is opposite to the inductance case. In that case, a larger inductance and a high frequency means less current. The reasons for the opposite results is not a mystery. A little careful thinking and arguing with one's self and reviewing the facts about capacitors will provide understanding.

If capacitance is bigger, the ingoing and outgoing currents are larger provided the applied voltage and frequency are the same. Hence the $\frac{1}{2\pi fC}$ rather than $2\pi fC$.

So then one concludes that in a capacitive circuit, the overall A. C. current is given by formula

$$I = \frac{E}{\sqrt{R^2 + X_C^2}} = \frac{E}{\sqrt{R^2 + \left(\frac{1}{2\pi fC}\right)^2}} = \frac{E}{\sqrt{R^2 + \left(\frac{1}{C\omega}\right)^2}}$$

If a circuit contains both inductance and capacitance, then one comes up readily and simply with the formula for overall A. C. current in another abridged form of Ohm's Law

$$\frac{E}{\sqrt{R^2 + \left(2\pi fL - \frac{1}{2\pi fC}\right)^2}} = I = \frac{E}{\sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2}}$$

This formula looks rather formidable but clear and careful thinking and re-reading of previous pages provides good understanding. The whole deal makes sense and is arrived at by simple arguments in step by step fashion.

CONCLUSIONS ABOUT L , C , R , AND f

One important fact is to be stressed and ever kept in mind. Energy stored in the magnetic field associated with the inductive reactance $2\pi fL$ (ohms) is not lost as heat but is stored energy and is recoverable. The energy stored in the electric field is the dielectric of a capacitor and associated with the capacitive reactance $\frac{1}{2\pi fC}$ is not lost as heat but is stored energy and is recoverable. Resistance energy is heat and is irrevocably lost.

These facts and ideas about energy stored in magnetic field around an inductance and in the electric field in the dielectric between plates of a capacitor and the recoverability of that energy will stand one in good stead in the study of oscillators and frequency of oscillation in radio and carrier systems. In these cases inductances and capacitors form a circuit sometimes called a tank circuit, an energy reservoir circuit, a take and give circuit.

If $C = 40$ mfd, $L = .176$ henries, then $f = 60$ cycles per sec.

If $C = 40$ mfd, $L = 9 \times .176$ henries, then $f = 20$ cycles per sec.

If $C = 40$ mfd, $L = 16 \times .176$ henries, then $f = 15$ cycles per sec.

When a 60 cycle resonance circuit is demonstrated in a laboratory the interrelations of resistance, inductance, and capacitance are nicely shown by the above equation or formula. The above formula gives interesting and important quantitative facts in an explicit and "short-hand" fashion. If a two electrode neon lamp is put in the tank circuit, interesting phenomena are convincingly portrayed. The neon lamp can be looked at in a tilting or a revolving mirror to reveal the alternations. A series of individual and successive flashes of one plate of neon lamp and then the other plate is observed in the mirror in a spread out line of alternate flashes.

RESONANCE AND OSCILLATION

It can be shown readily by an experiment in the laboratory that the so called tank circuit of inductance and capacitance resonates to a 60 cycle per second input.

In other words, there was a resonant interchange of the energy stored in the capacitor reservoir and the energy stored in the inductance reservoir. From this standpoint, the inductance reactance and capacitive reactance were equal to one another and offset one another as it were, so that the A. C. supply thought that there was only resistance in the circuit.

Therefore, let us take the $2\pi fL$ and $\frac{1}{2\pi fC}$ from the above formula and put them equal to one another

$$2\pi fL = \frac{1}{2\pi fC}$$

By some simple algebraic maneuvers one arrives at a very interesting and very important relation

$$f = \frac{1}{2\pi \sqrt{LC}}$$

$$f = \frac{\text{cycles}}{\text{sec}} = \frac{1}{2\pi \sqrt{LC}} = \frac{.159}{\sqrt{LC}}$$

In the tank circuit, there is an interchange of energy. It is to be remembered that the capacitor energy is recoverable and likewise the inductor energy. They give and take one another's energy back and forth, forth

and back, on and on in time. One is reminded of water stored initially in two tanks at different levels (voltage) flowing back and forth through a connecting pipe when a valve in the pipe is opened. As the circuit oscillates, little energy is lost in the oscillating water circuit itself. If a bit of energy in the reservoirs of the electrical tank circuit is lost in the resistance of the wires, the 60 cycle A. C. will supply a new bit of energy and keep "nudging" the tank oscillations so that they will be maintained. In an oscillating circuit with a vacuum tube as a component part and a *direct* voltage supply, the vacuum tube plays the role of the "nudger" at just the right time to maintain oscillation.

One can think of a child's swing which can be maintained by a little "push", a little nudge, at just the right instant of time by another child standing by.

If one looks into matters a bit further, it will be found without too many mental fogs reducing the visibility, that the amount of resistance R , which will not take away too much energy in terms of wasted heat to permit the capacitive and inductive energy to interchange and oscillate, is given by a simple formula.

If R is less than $2\sqrt{\frac{L}{C}}$ then oscillations will occur and be maintained. In this case the L will be in henries, the C in farads and the R in ohms.

INTERPRETATION OF FORMULA $2\sqrt{\frac{L}{C}}$

This $2\sqrt{\frac{L}{C}}$ again illustrates how explicit and valuable information to help in design and operation is provided by a simple formula. Interpreting this formula, one concludes that if one has a choice among different L 's and different C 's for the oscillating circuit to get a given frequency as per formula $\frac{1}{2\pi\sqrt{LC}}$, it is better from the standpoint of $2\sqrt{\frac{L}{C}}$, to select a small C and a large L . The $\frac{1}{2\pi\sqrt{LC}}$ will be kept the same, but $2\sqrt{\frac{L}{C}}$ will become greater. If it is greater, then the R in the circuit will have a better chance to be less than the $2\sqrt{\frac{L}{C}}$.

As previously discussed in the section on capacitance in this memorandum, the amperes into a capacitor is larger when the capacitance (microfarads) is larger. On a previous page this sentence appears.

"This means that if C is bigger and f higher, the overall value of current into C will be bigger."

So if by choosing a small C and large L , the effective A. C. current in the oscillating circuit will be less. Since the heat loss in a circuit is I^2R , there is less loss in the circuit and more likelihood of sustained oscillations.

These facts are determinable because the formula $2\sqrt{\frac{L}{C}}$ gives us the clue.

In the old days of radio receivers with separate dials for L and C , one could "get" an incoming signal with various settings of the L and C dials. Availing oneself of the $2\sqrt{\frac{L}{C}}$ idea, the signal would be better if the L dial were at a large setting and the C dial at a small setting. This meant as per above argument of less A. V. current in the tuned circuit and less absorption of energy (heat I^2Rt) and there would be more likelihood of a better signal.

Of course if L is larger, there is more chance of a greater R due to more wire with some resistance in the larger L . But heat loss is I^2Rt , so if current is smaller due to smaller capacitor, there is an overall advantage due to the current square argument in spite of the larger R in the larger L itself. The origin and background of the $2\sqrt{\frac{L}{C}}$ criterion is discussed in the next pages.

CHAPTER 10

Lord Kelvin's Equation for Electrical Oscillations and Damped Oscillating Current.

In the discussion in pages 115-118, (Figures 46, 47), the equations of damped waves were presented in general mathematical terms. It was pointed out that the alternating current damped amplitude curve embodied two distinct parts. The curve can be regarded as the result of the product of simultaneously occurring values of two curves. One of the curves is a sine wave curve per se and the other a decreasing exponential curve of geometric retrogression nature. The sine curve has the equation

$$I_1 = 20 \sin \theta$$

The exponential curve has the equation

$$I_2 = 20 \cdot e^{-.02\theta} = 20 \cdot .98^\theta$$

The combined curve equation of the damped wave is

$$I_3 = 20 \sin \theta e^{-.02\theta} = 20 \sin \theta \cdot .98^\theta$$

As stated above, these are generalized or implicit equations and do not embody explicitly the three basic factors of an electrical circuit namely, resistance, inductance and capacitance.

In 1854, Sir William Thomson, Lord Kelvin first published the complete analysis for an oscillating electrical circuit containing R , L and C . The differential equation was provided by invoking Kirchoff's Law. This basic and simple law states that when a capacitance is energized or charged and then short-circuited and thereby the discharge is effected through the resistance and inductance, the sum of the three voltages involved at every instant of time is equal to zero.

Kirchoff's Law expressed in the symbols of mathematics is as follows.

$$L \frac{di}{dt} + Ri + \int \frac{i dt}{c} = 0 \quad \text{Here } i dt = dq$$

Inductance voltage + resistance voltage + capacitance voltage = 0.

The consummate mathematical skill of Kelvin combined with a keen insight into the basic electrical principles and concepts enabled him to solve the differential equation and provide an explicit equation for i , the electric current at any and every instant of time.

$$i = \frac{Q}{LC \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}} e^{-\frac{R}{2L}t} \sin \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} t.$$

Q is the maximum initial electrical charge (coulombs) in the energized capacitor and equals CE , where E is the energizing applied voltage and C the capacitance in farads.

During the oscillations, the magnetic field energy embodied in the space around the inductance and the electric field energy embodied in the dielectric between the plates of the capacitance interchanges at a frequency prescribed by R , L and C . Also during the oscillations, energy is irrevocably lost in the heat radiated by the resistance. This latter energy loss takes place in accordance with the well known formula, $i^2 R t$.

* For an excellent and complete discussion of the derivation of this formula see Bedell and Crehore "Alternating Currents" 5th edition, 1923.

The above explicit expression for the instantaneous current can be compared to generalized equations discussed in pages 115, 118 Figures 46, 47. The comparison of the explicit and implicit equations afford illuminating ideas and understanding.

The generalized or implicit equation is

$$i = I_0 e^{-\kappa t} \sin \theta$$

Since $\theta = \omega t$ where ω (omega) is the angular velocity one writes $i = I_0 e^{-\kappa t} \sin \omega t$. Furthermore since ω is $2\pi f$ where f is the frequency of oscillation, one writes $i = I_0 e^{-\kappa t} \sin 2\pi f t$.

Comparing part by part, the explicit equation of Kelvin with the implicit equation

$$I_0 = \frac{Q}{LC \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}}$$

As an approximation $I_0 = \omega Q = \omega CE$. This approximation will be discussed in detail later in this chapter.

This approximation is mentioned now because it shows that the smaller the C , the smaller the initial value of I_0 at time, $t = 0$. At all subsequent instants of time, the current will also of course be smaller, the smaller the C . Since heat loss is proportional to i^2 , it well for continuance of oscillations to have small current to lessen the rate of heat loss.

Further comparison of the two types of equations reveals at a glance that

$$\omega = 2\pi f = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

This observation provides the clue for an expression or formula for the frequency of oscillation. If one assumes that the resistance R in ohms is small then the second part of the formula under the square root sign $\frac{R^2}{4L^2}$ will be small and approach zero and therefore be neglected. Of course R can never be zero so the formula resulting from the assumption that $\frac{R^2}{4L^2}$ is zero will be an approximate formula.

One can write therefore

$$2\pi f = \frac{1}{\sqrt{LC}}$$

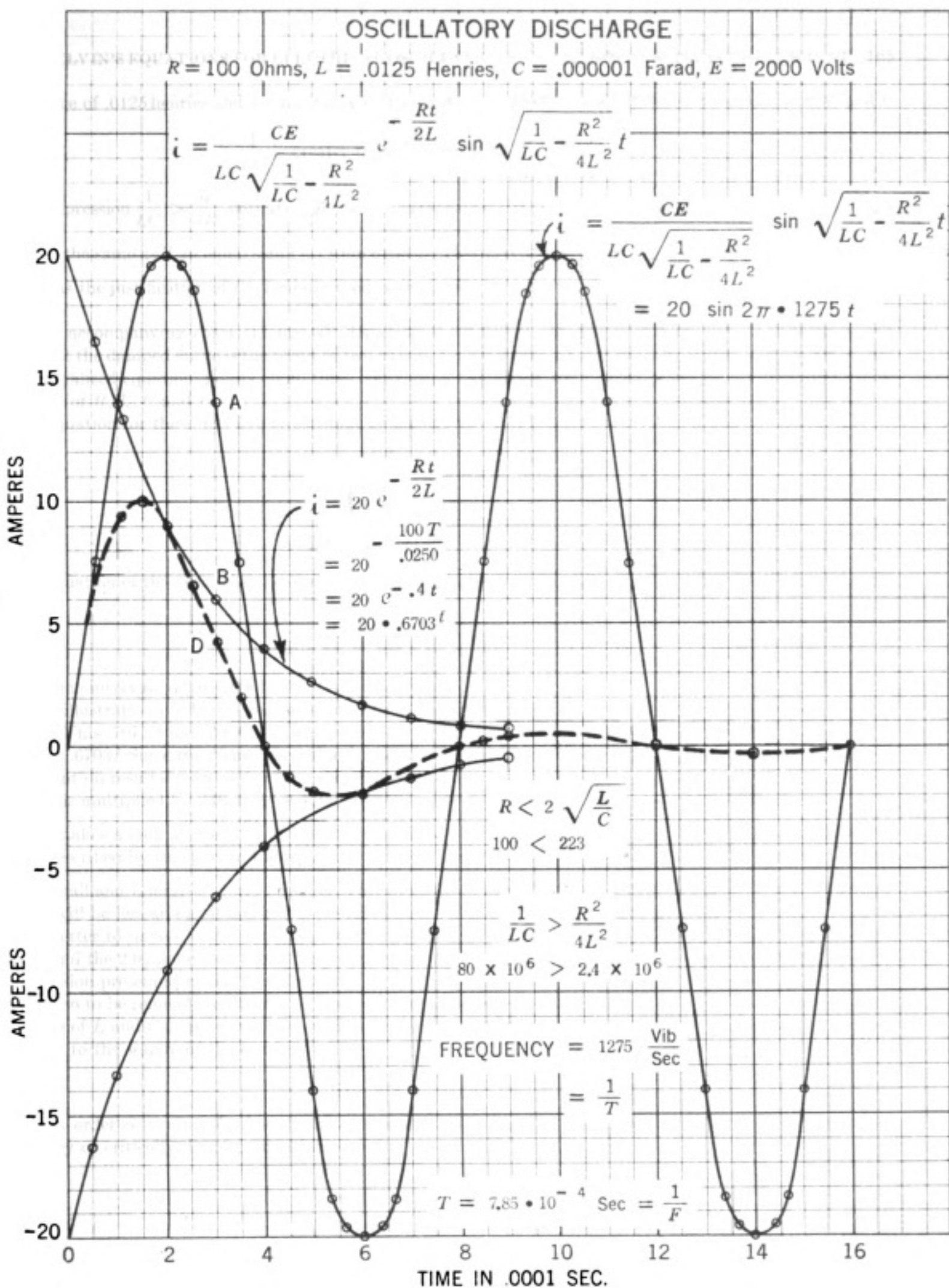
$$f = \frac{1}{2\pi\sqrt{LC}}$$

The period of one complete oscillation $T = \frac{1}{f} = 2\pi\sqrt{LC}$. This approximate formula for frequency has been used ever since 1854 by the communication engineer to study, experiment and explore the nature and frequency of an electrical oscillatory circuit. As said before, it is not possible to have a circuit with zero resistance but the resistance can be very small to provide validity to the approximate formula.

Furthermore the explicit formula in terms of R , L and C compared to the implicit formula $\omega = 2\pi f = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$, provides the clue as to whether or not, in the interchange of energy of magnetic field ($\frac{1}{2}Li^2$) and the electric field energy $\frac{1}{2}Ce^2$ (e = instantaneous voltage) and the loss of energy in i^2Rt , the conditions are such that there will be alternating current oscillations. It might be that the discharge would be non-oscillatory and there would be a "dead beat" collapse; a single rise to a maximum and then a fall to zero of direct current. Take a preliminary look at Figures 60 and 61.

For ω or $2\pi f$ to be a positive and real number, or in other words for oscillations to occur, it is seen by inspection that the R , L and C under the square root sign must be so proportioned that $\frac{1}{LC}$ be greater than $\frac{R^2}{4L^2}$.

The accompanying chart (Figure 60) portrays an oscillating current in a circuit with an R of 100 ohms, an



inductance of .0125 henries and a capacitance of 1 microfarad (.000001 farad). With these values of R , L and C

$$\frac{1}{LC} \text{ is greater than } \frac{R^2}{4L^2}: 80 \times 10^6 > 2.4 \times 10^6$$

The expression $\frac{1}{LC} > \frac{R^2}{4L^2}$ permits algebraic re-arrangement of terms with respect to its greater than sign ($>$) just the same as if an equal sign ($=$) were present. Therefore R is less than $2\sqrt{\frac{L}{C}}$; $R < 2\sqrt{\frac{L}{C}}$. This expression is the justification of the formula used in page 161. With the values of R , L and C of this circuit $100 < 223.6$.

In the accompanying chart (Figure 60) depicting an oscillatory discharge current, it is to be pointed out again that the damped curve is the result of two curves. In this present case of an explicit equation as contrasted to a generalized equation, points on all three curves were calculated and plotted by using the numerical values assumed for R , L , C and E .

The equation for the sine wave of constant amplitude is

$$i = \frac{CE}{LC \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}} \sin \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} t$$

$$= 20 \sin 2\pi \cdot 1275t$$

The equation of the exponential curve which decrees a geometric retrogression reduction of values of the sine curve is

$$i = 20e^{-\frac{R}{2L}t} = 20e^{-\frac{100}{.025}t} = 20e^{-.4t} = 20 \cdot .6703^t$$

The time intervals (t) are .0001 (10^{-4}) seconds.

As an illustration of the calculations, consider a time interval of 3 units (.0003 seconds). The sine of 135° is .7071. This .7071 times 20 = 14.142 amperes. This point on curve is marked A . The value of $e^{-.4 \cdot 3}$ or $e^{-1.2}$ is .301 or .6703.³ Since the 20 ampere value for I_0 has been used once, the .301 is not multiplied by 20 again to get a point on resultant curve. On the exponential curve per se, the point in question is marked B . The 14.142 amperes is multiplied by .301 to give 4.25 amperes point on the damped curve which is marked D .

If one takes a look at the $e^{-\frac{R}{2L}t}$ in Kelvin's equation, it is noted that the enervating geometric retrogression effect prescribed by this part of the equation on the pure sine wave is less if R is small and inductance is large.

If R is small and L large $e^{-\frac{R}{2L}t}$ will be small. This means of course that the constant loss geometric ratio per unit of time will be less and a situation will prevail which is conducive to an oscillatory result.

It is better to have $i = I_0 e^{-.01t} = I_0 \cdot .9900^t$ than to have $i = I_0 e^{-2.01t} = I_0 \cdot .1340^t$. It will be recalled that the .01 and the 2.01 are called logarithmic decrement constants (the K 's) in the general discussion of geometric retrogression presented in the Introduction.

It is also to be remembered that a relatively low capacitance means less current in the various possible combinations of L and C to produce the same frequency. The less current means less $i^2 R t$ loss since the loss is proportional to the square of the current.

QUALITY OF A CIRCUIT, Q

It is in order to analyze further the matter of oscillations and critical damping. The criterion for the non-oscillatory or critically damped situation is that

$$\frac{1}{LC} = \frac{R^2}{4L^2}$$

We know the valid approximate expression for frequency, $f = \frac{1}{2\pi\sqrt{LC}}$ and that the period $T = \frac{1}{f} = 2\pi\sqrt{LC}$.

$$\omega = \text{angular velocity} = 2\pi f = \frac{1}{\sqrt{LC}}; \omega^2 = \frac{1}{LC}$$

From above equation

$$\frac{1}{LC} = \frac{R^2}{4L^2} = \omega^2$$

In turn

$$\frac{L^2\omega^2}{R^2} = \frac{1}{4}; \quad \frac{L\omega}{R} = \frac{1}{2}$$

For identification and description, one calls the $\frac{L\omega}{R}$, the quality factor (Q) of a circuit. In other words $Q = \frac{L\omega}{R}$ by definition. This Q must not be confused with electrical charge (coulombs) in which case $Q = \text{Voltage times Capacitance}$ ($Q = CE$). The present Q is the ratio between $L\omega$ (inductance reactance, X_L) and resistance, R .

In terms of Q , the quality of a circuit, it is to be said that the quality of a circuit must be greater than $\frac{1}{2}$ in order that the circuit be oscillatory.

If the frequency is assumed to be $\frac{1}{2\pi\sqrt{LC}}$, then the period $T = \frac{1}{f} = 2\pi\sqrt{LC}$. Now T also $= \frac{2\pi}{\omega}$. The ratio of successive positive maximum (or negative) amplitudes of a damped oscillatory wave train is decreed by the damping exponential curve $i = Ie^{-\alpha t}$ where $\alpha = \frac{R}{2L}$ and the t is now T , the period of oscillation. The ratio in question is

$$r = e^{-\alpha \cdot T} = e^{-\frac{R}{2L} \cdot \frac{2\pi}{\omega}} = e^{-\frac{\pi R}{L\omega}} = \frac{\pi R}{e^{L\omega}}$$

LOGARITHMIC DECREMENT AND QUALITY

As previously discussed in pages 181-184, the logarithmic decrement is defined as the logarithm to the base e of the ratio between two positive (or negative) maximum amplitudes. The logarithmic decrement is designated as δ (delta).

The logarithm of " r " is the logarithm of $e^{\frac{\pi R}{L\omega}}$ and

$$\log_e \frac{\pi R}{e^{L\omega}} \text{ is } \frac{\pi R}{L\omega}; \quad \delta = \frac{R}{\pi L\omega}$$

In the discussion and numerical examples in pages 119-123, the logarithmic decrement was an implicit quantity. Now we have an explicit expression for δ in terms of the electrical factors resistance and inductive reactance.

Since $\omega = \frac{1}{\sqrt{LC}}$, one can write $\delta = \pi R \sqrt{\frac{C}{L}}$. Now we have an explicit expression for δ in terms of R , L and C .

The quality of a circuit was defined as $\frac{L\omega}{R}$. Therefore $\delta = \frac{\pi}{Q}$. It follows that $Q = \frac{1}{R\sqrt{\frac{C}{L}}}$. This is an explicit

expression for Q in terms of R , L and C . It can also be written $Q = \frac{\sqrt{L}}{R\sqrt{C}}$

In a particular circuit with known values for these three electrical quantities, one can calculate and predict in advance as to whether or not, the circuit would be an oscillatory one, on the basis that the Q must be greater than $\frac{1}{2}$.

The formula for $Q = \frac{1}{R\sqrt{\frac{C}{L}}} = \frac{\sqrt{\frac{L}{C}}}{R}$ could have been derived readily by a different approach.

The criterion for oscillation as previously stated is $\frac{1}{LC}$ must be greater $\frac{R^2}{4L^2} \cdot \frac{1}{LC} > \frac{R^2}{4L^2}$ Rearranging the terms in this last expression by simple algebraic procedures

$$\frac{4L^2}{LC} > R^2: \quad 2\sqrt{\frac{L}{C}} > R: \quad \frac{\sqrt{\frac{L}{C}}}{R} > \frac{1}{2}$$

Q was defined as $\frac{L\omega}{R}$. It can also be defined as $\frac{\sqrt{\frac{L}{C}}}{R}$. The two definitions are intrinsically the same but the second definition is a bit preferable since it expresses Q in terms of R , L and C . The approach and derivation immediately above shows quickly and convincingly that Q must be greater than $\frac{1}{2}$ in order that there be oscillations.

Consider the oscillatory circuit depicted in the accompanying chart Figure 60. The circuit constants are $R = 100$ ohms, $L = .0125$ henries and the $C = .000001$ farads. The calculated $Q = 1.1$ and the frequency is 1275 cycles per second.

On page 161, the test or criterion formula was that R must be less than $2\sqrt{\frac{L}{C}}$. In the present case $100 < 223$ We now have the origin and validation of the formula

$$Q = \frac{\sqrt{\frac{L}{C}}}{R} > \frac{1}{2} \quad \text{which is to say} \quad R < 2\sqrt{\frac{L}{C}}$$

RELATION OF Q AND δ

It was shown that $\delta = \frac{\pi}{Q}$. If for a critically damped circuit $Q = \frac{1}{2}$ then in turn $\delta = 2\pi$ (6.28). If Q is to be greater than $\frac{1}{2}$ for an oscillatory circuit, then the logarithmic decrement must be less than 6.28.

The assumption in pages 119-123 that the oscillatory circuit has petered out and the oscillations can be regarded as negligible when a positive amplitude toward the end was .01 of the initial positive amplitude was only a convenient assumption. In the illustrations previously given, the number of oscillations was given by

the formula $n = \frac{\log 100}{\delta}$. When δ was .2 which meant a ratio of 1.22 for two successive positive or negative

amplitudes, the $n = \frac{\log 100}{\log 1.22} = \frac{4.6}{.2} = 23$ oscillations. The .01 (1%) value would be the ratio after 4.6 time

constants where the time constant is $\frac{2L}{R}$. It might be better in these present days of sensitive receivers to assume

a ratio of .001 or .1%. 1% is a 20 decibel loss and .1% is a 30 decibel loss. A value of .001 is decreed by the exponential curve after approximately 7 time constants. If this were the assumption, the numbers of oscillations

n would be $\frac{\log 1000}{\log 1.22} = \frac{7}{.2} = 35$ oscillations. All logarithms are to the base e

In terms of the fact that δ must be less than 6.28, one can find the approximate number of oscillations in a minimum situation when the minimum of assumed negligibility is .001 after 7 time constants. Again as an assumption, assume a maximum of 6 for δ . In turn this means since $Q = \frac{\pi}{\delta}$, the value of Q of the circuit would be .526 which is not much greater than .5 to be sure but it is greater than .5

If δ were 6, the ratio of two successive positive amplitudes would be 403, since $e^6 = 403$. This is indeed a very large ratio and a very high damping. The wave train would peter out very quickly to negligibility unless there was an immense amount of energy put in the system originally by a very high voltage supply.

With assumed negligibility of .001 and an assumed δ of 6, $n = \frac{\log 1000}{\log 403} = \frac{7}{6} = 1.16$ vibrations. This 1.16 oscillations is not very many. However the analysis has shown how Q , δ and " n " are related. The upshot of the matter is that with the seeming reasonable assumption of .001 for negligibility, one must have a Q which is much greater than $\frac{1}{2}$. If the assumed negligibility (.001) is reasonable, then to have some semblance of continued oscillations say 500, the δ must be $\frac{\log 1000}{500} = \frac{7}{500} = .014 = \delta$. A value of .014 for δ means the ratio of two successive amplitudes would be 1.014. The assumption of 500 oscillations being a reasonable number means in turn that the Q of the circuit would be 224.4 since $Q = \frac{\pi}{\delta} = \frac{\pi}{.014} = 224.4$.

Analyzing in reverse at it were, the number of oscillations on the assumption of .001 negligibility for a circuit with a value of $Q \left(\frac{1}{R\sqrt{\frac{C}{L}}} \right)$ as determined by its value of R , L and C , is given by the formula $n = \frac{7Q}{\pi}$.

If one had a circuit whose Q was 100, n would be 223. The δ for this assumed circuit would be .31416 and and " r " would be 1.36

The Q of the oscillatory circuit depicted in Figure 60 was calculated as 1.1 on a previous page. The " n " as shown above is $\frac{7Q}{\pi}$ which for this circuit would $\frac{7.7}{\pi}$ or 2.45 complete oscillations which checks quite well with the sketch remembering the .1 % negligibility assumption.

APPROXIMATE FORMULAS

Actually the preceding formulas are approximate ones. Of course Kelvin's basic equation is not approximate. The formula for frequency $\left(\frac{1}{2\pi\sqrt{LC}} \right)$ is valid only on the basis that R^2 is very, very small compared to $4L^2$. The exact formula for frequency as given by Kelvin's equation is $\frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$. The frequency of the oscillatory circuit (Fig. 60) where $R = 100$ ohms, $L = .0125$ henries and $C = 1$ microfarad is 1275 cycles per second. The approximate frequency on the basis of the formula $\frac{1}{2\pi\sqrt{LC}}$ is 1424 vibrations per second.

Furthermore, the initial current is 20 amperes on the basis of Kelvin's equation in which $I = \frac{Q}{LC \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}}$. On a previous page, it was stated that the initial current is approximately equal to ωCE .

The $\omega = 2\pi f = 2\pi \cdot 1424$. The C is .000001 farads and the E is 2000 volts. By the approximate formula, the initial current is 17.4 amperes.

On the first page of the present discussion of quality of circuit, (Q) it was shown that on the basis that $f = \frac{1}{2\pi\sqrt{LC}}$, that ω^2 equaled $\frac{1}{LC} = \frac{R^2}{4L^2}$ or $\omega = \sqrt{\frac{1}{LC}} = \frac{1}{2} \sqrt{\frac{R}{L}}$ or $2\pi \cdot 1424$ or 8,744 radians per seconds. Actually ω is neither of the above. In the circuit concerned, the ω is $2\pi \cdot 1275$ or 7,828.5 radians per second.

As per Kelvin's equation

$$\begin{aligned}\omega_0 &= \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \\ &= 2\pi f_0 = 2\pi \cdot 1275 \\ &= 7,828.5\end{aligned}$$

The ω_0 can be written $= \omega \sqrt{1 - \frac{1}{4Q^2}}$

So that ω be equal to ω_0 (the real angular velocity) and all factors be very nearly correct, the Q must be large so that $\frac{1}{4Q^2}$ approach zero.

The upshot of the situation is that even the formula for Q is approximate and the expression for $\delta = \frac{Q}{\pi}$ is approximate and the formula $n = \frac{7Q}{\pi}$ is approximate.

In other words, if one wishes to use these helpful and approximate formulas to know about a circuit, the δ 's must be small and the Q 's must be large.

SPARK RADIO TELEGRAPHY

In the yesterdays when radio-telegraph spark systems were the vogue, a specification for such systems was that the logarithmic decrement should not be greater than .2. As shown before, this means a ratio of 1.22 for successive positive or negative amplitudes in the wave train. Logarithmic decrement is obviously associated with rate of damping and sharpness of resonance or sharpness of tuning. If a wave train of a radio telegraph spark system had a δ greater than .2, the tuning to receive the signal was broad; it could be heard on quite a range of the tuning dials. In those yesteryears an instrument called a decremeter was used as a testing instrument to appraise the wave train output of the transmitter. The Kolster decremeter was a valuable instrument

If the logarithmic decrement were .2 this meant that circuit's Q was $\frac{\pi}{.2}$ or 15.7. On the basis of assumed negligibility of 30 debibels loss or .001, there were 35 complete oscillations in each wave train.

Finally it is in order to say that the approximate formulas for Q , δ , and n in terms of R , L and C were evolved from simple and fundamental truths combined with basic mathematical relationships. The explicit formulas convincingly illustrate the fact that if an oscillatory circuit is to perform acceptably in a communication system, it must have a sizable Q . The sizable Q results from the fact that R is very small, that L is relatively and that C is relatively small.

JUST NON-OSCILLATORY DISCHARGE; DIRECT CURRENT DISCHARGE

If $\frac{1}{LC}$ is equal to $\frac{R^2}{4L^2}$, one has a zero under the square root sign of Kelvin's equation. This might mean that there was no discharge at all and of course this could not be. If $\frac{1}{LC}$ were smaller than $\frac{R^2}{4L^2}$, then one would have a negative quantity under the square root sign. Both situations would be a bit disturbing. So the engineer who is well acquainted with the value and importance of mathematical understanding and analysis in resolving problems of a practical nature is confronted with a dilemma. Having ever been grateful to the skilled and expert mathematician for guidance and help, the engineer again seeks assistance. The engineer has his talent, his skill and know-how but he does not have knowledge of mathematical methods and techniques per se to resolve the situation when the square root part for angular velocity in the explicit Kelvin equation is either zero or negative.

For the case when $\frac{1}{LC} = \frac{R^2}{4L^2}$, his problem is solved by accepting with thanks a new equation which mathematics per se provides.

The equation applicable to the just non-oscillatory situation which can be called the critically damped case is as follows

$$i = \frac{E}{L} t \cdot e^{-\frac{R}{2L}t}$$

Examination of this new equation reveals, that the varying direct current occurring on discharge of capacitance is the product of two factors. One factor is the straight line relation $\frac{E}{L} t$ which prescribes an arithmetical

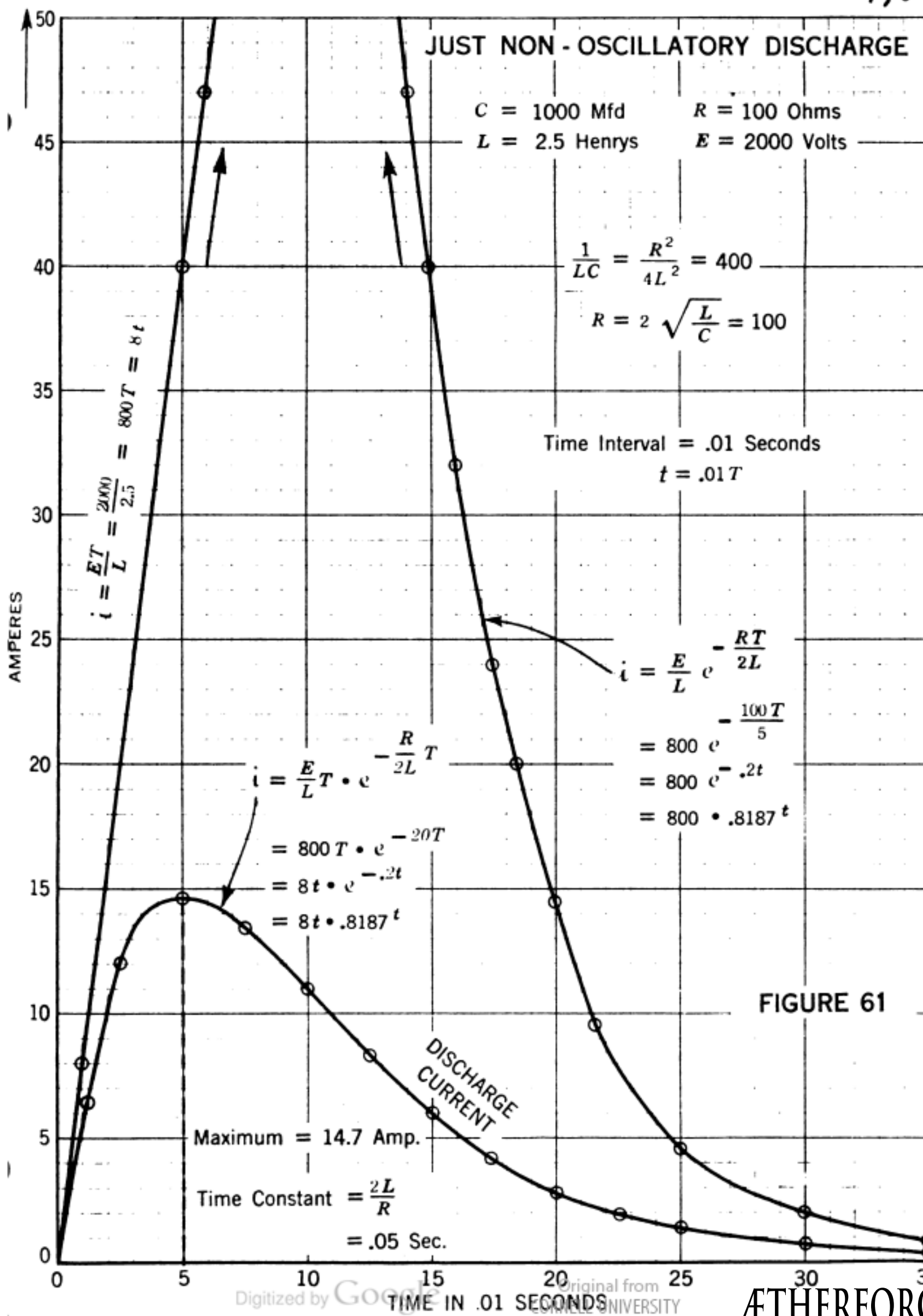


FIGURE 61

progressive relation between i and t . The slope of the straight line is $\frac{E}{L}$. The second factor is the geometric retrogression factor $e^{-\frac{R}{2L}t}$. In the case of the damped oscillatory current, there was a sine wave and an exponential factor which combined to give the resultant current

But now, by way of interesting contrast, one of the three curves involved is a straight line.

The accompanying chart, Figure 61, depicts the just non-oscillatory case. The resistance is 100 ohms, the C is .00100 farads (1000 microfarads) the L is 2.5 henries and the voltage is 2000 volts Using these values

$$\frac{1}{LC} = \frac{R^2}{4L^2} = 400$$

CALCULATION OF Q , QUALITY OF CIRCUIT

From the standpoint of Q as discussed in immediately preceding pages, it is to be expected that $Q = .5$, the value of Q for critical damping

$$Q = \frac{1}{R\sqrt{\frac{C}{L}}} = \frac{1}{100\sqrt{\frac{.001}{2.5}}} = .5$$

In a circuit of .00001 farads, 100 ohms and .0016 henries, the Q is .4; this would be an "over" damped circuit.

CALCULATION OF CURRENT PORTRAYED IN FIGURE 61.

If the time in seconds is designated by T and the time is then expressed in units of .01 second as t , one can write

$$i = \frac{E}{L} T = \frac{2000}{2.5} T = 8t$$

in the straight line. The critically damped current is then

$$i = \frac{E}{L} t \cdot e^{-\frac{Rt}{2L}} = 800t \cdot e^{-.2t} = 800t \cdot .819^t$$

TABLE

t (.01 Seconds)	i (Straight Line)	$e^{-.2t}$	i , D.C. Discharge
	<i>amperes</i>		<i>amperes</i>
1.0	8	$e^{-.2} = .819$	6.5
2.5	20	$e^{-.5} = .606$	12.1
5.0	40	$e^{-1.0} = .368$	14.7
7.5	60	$e^{-1.5} = .223$	13.4
10.0	80	$e^{-2.0} = .135(.368)^2$	10.8
12.5	100	$e^{-2.5} = .082$	8.2
15.0	120	$e^{-3.0} = .050(.368)^3$	6.0
17.5	140	$e^{-3.5} = .030$	4.2
20.0	160	$e^{-4.0} = .018(.368)^4$	2.8
22.5	180	$e^{-4.5} = .011$	2.0
25.0	200	$e^{-5.0} = .0067(.368)^5$	1.3
30.0	240	$e^{-6.0} = .0025(.368)^6$	0.6
35.0	280	$e^{-7.0} = .0009(.368)^7$	0.25
100.0 (1 second)	800	$e^{-20.0} = .0000000$	0.0000000

It is to be observed that the values of i in last column are the product of values in second and third columns.

The time constant concept is discussed at length in later pages (180-219) pertaining to transient phenomena in electrical circuits. It can be said now that the time constant is the time necessary to make the negative

exponent of e be unity. It is seen that when $t = \frac{2L}{R}$, $e^{-\frac{R}{2L} \cdot \frac{2L}{R}}$ would be e^{-1} . At .05 seconds or 5 of the .01 second intervals, the D.C. current reached its maximum value of 14.7 amperes as is shown in Figure 61. Therefore the time-constant in this situation (.05 seconds) is defined as the time elapsing until the rising and falling direct current discharge reaches maximum value.

After an elapsed time of 7 time constants, namely .35 seconds, the current is .0009 times 280 amperes or .25 amperes. This .0009 is very near to .001 and provides the basis for the criterion of negligibility. If the .001 value is assumed it would mean a 30 decibel loss. This would give .28 amperes which is a big greater but nevertheless the 7 time constant assumption for negligibility is a valid one.

ARITHMETICAL PROGRESSION AND GEOMETRIC RETROGRESSION

The uni-directional current discharge curve of Fig. 61 has a most interesting origin and make up. The value of current of each point along its contour as a function of time is a product and obtained by the multiplication of values of points on the other two curves. These other two curves are widely different in nature. One of the curves is a straight line which portrays an increase of direct current a la arithmetical progression fashion as time increases. The second is the exponential curve which portrays a decrease a la geometric retrogression fashion as time increases. In the end, the second curve of rapid decrease overwhelms the first curve with its rapid increase so the resultant current peters out to zero. The combination provides a rather nondescript curve of no symmetry mathematically speaking but from a physical and electrical point of view tells an interesting story.

The above situation is an interesting contrast to the case where a sine curve and an exponential curve combine to give a damped oscillatory wave.

In later pages 354, Figure 90, when the hyperbolic functions are discussed and depicted by curves, it will be shown that an e^x geometric progression curve is the sum of two curves, namely $e^x = \cosh x + \sinh x$. The e^{-x} geometric retrogression curve is the difference of two curves, namely $e^{-x} = \cosh x - \sinh x$.

OSCILLATING CIRCUIT EXPERIMENT INDUCTANCE AND CAPACITANCE

The accompanying sketch shows (Figure 62) a simple circuit arrangement.

A capacitor, 40 microfarads, is charged to a voltage of 100 or 200 volts. It will be remembered that in a previous experiment described in this memorandum, when a capacitor was charged and discharged through a resistance of 2,000 ohms, only one disc of the neon lamp voltmeter glowed on charge, stayed lighted and the same disc glowed for a while on discharge. In this present case, the microfarad condenser is first to be charged and then allowed to discharge through an inductance of 7.03 henries and low ohmic resistance by the closing of the #2 switch. In this latter case, both disks of the neon lamp will alternately glow for a little while, showing the discharge is oscillatory. The current in the oscillatory circuit with 60 cycles per second changes in direction so the polarity of the voltage across the neon lamp changes. The discs alternately glow but they do so, so rapidly that the naked eye cannot see the alternations. However, if the neon lamp is look at in a small plane mirror, the alternate glows can nicely be observed when the mirror is quickly tilted back and forth.

If a bit larger capacitance and a considerably larger inductance are used so that the frequency of oscillation according to the formula $f = \frac{1}{2\pi\sqrt{LC}}$ is about 20 or 15 cycles per second, the naked eye can observe the separate and alternate glowings of the two neon lamp discs. One wants to remember the $2\sqrt{\frac{L}{C}}$ idea so that the oscillations will be continued for quite a while.

In Canada many of the commercial A.C. power systems operate at 25 cycles. In this case, the naked eye cannot quite observe the variations in brightness of an incandescent lamp nor could the 25 cycles be observed by the naked eye if a neon lamp were used. But if a tilting mirror is used, then the alternation of 25 and 60 cycles too can beautifully be observed in the tilting or rotating mirror. Hence, in this suggested oscillatory case, the frequency out to be 20 or 15 cycles per second to be noticeable to the unaided eye. It is likely that a capacitance of 9 microfarads and an inductance of 49 henries would give about a frequency of about 8 cycles per second. Of course the pure ohmic resistance of the 49 henry coil would need to be low to get oscillations.

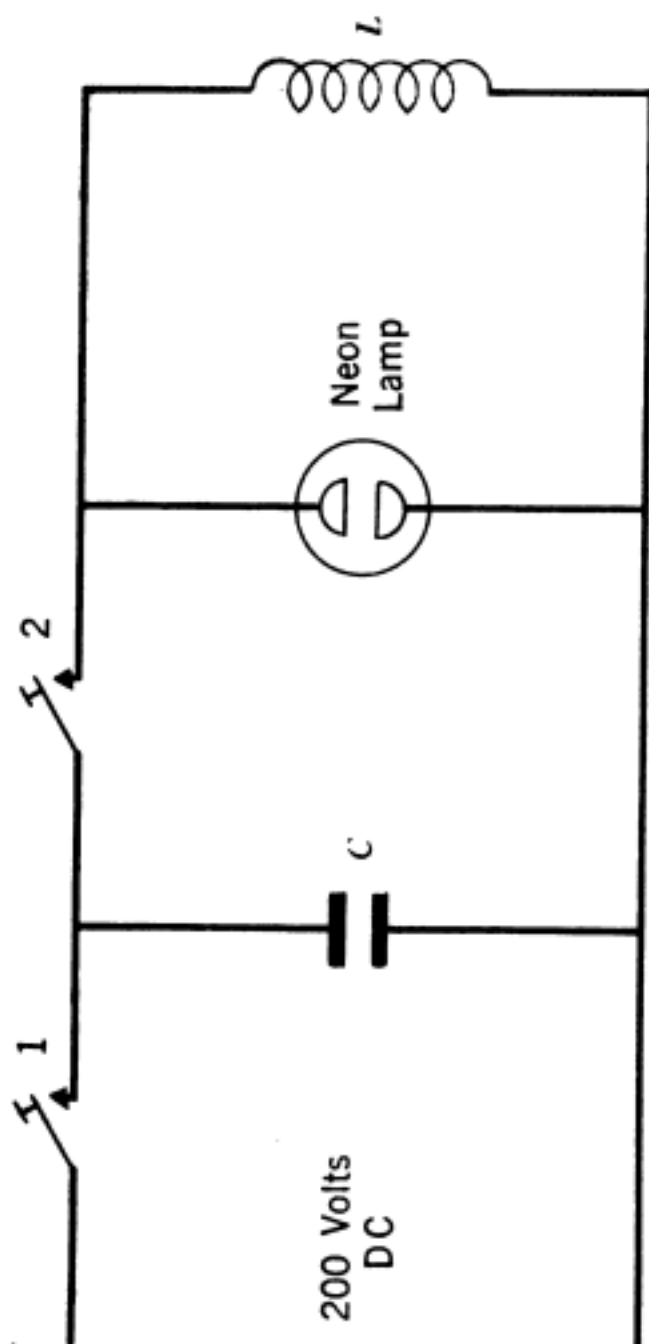


FIGURE 62

OSCILLATING CIRCUITS IN ELECTRICAL COMMUNICATIONS

Wherever the electrical oscillating circuit is found, 60 cycles power system, audio oscillators, heat frequency oscillators, high frequency oscillators in carrier and radio systems, this very same explicit equation for the frequency tells the story in shorthand fashion.

As has been said many times, resistance, (R), is a factor which translates electrical energy into non-recoverable heat. Hence, in the oscillating circuits in all kinds of electrical communications equipment, an all out effort is made to keep R low, below $2\sqrt{\frac{L}{C}}$. It is best to have C small and L large so $2\sqrt{\frac{L}{C}}$ may be big. The equation or formula for frequency of an electrical oscillating circuit is as given before,

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{.159}{\sqrt{LC}} \text{ where } L \text{ is henries and } C \text{ is farads.}$$

L , the inductance which stores and gives back its magnetic field energy, can be expressed in henries, millihenries or microhenries. C the capacitance which stores and gives back its electric field energy is usually expressed in microfarads. The two energies interchange.

If one puts L and C in different units, the above equation can be written in several ways.

$$\begin{aligned} \text{frequency} &= \text{cycles per second} \\ &= \frac{.1592}{\sqrt{L \text{ (henry)} C \text{ (farad)}}} \\ &= \frac{159.2}{\sqrt{L \text{ (henry)} C \text{ (microfarad)}}} \\ &= \frac{5033}{\sqrt{L \text{ (millihenry)} C \text{ (microfarad)}}} \\ &= \frac{159,200}{\sqrt{L \text{ (microhenry)} C \text{ (microfarad)}}} \\ &= \frac{159,200,000}{\sqrt{L \text{ (microhenry)} C \text{ (micromicrofarad)}}} \end{aligned}$$

Micromicrofarads are sometimes called picofarads.

If in radio one wishes to know wave length, then one must divide the velocity of electromagnetic waves by the frequency. Velocity is 3×10^8 meters per second; 186,000 miles per second.

$$\text{Wave Length} = \frac{3 \times 10^8 \text{ (meters per second)}}{\text{frequency (cycles per second)}}$$

After working the slide rule or doing the arithmetic of division one comes up with the following good formulas to have in one's notebook.

$$\begin{aligned} \text{Wave Length} &= \text{meters} \\ &= 1.884 \cdot 10^8 \sqrt{L \text{ (henry)} C \text{ (farad)}} \\ &= 1.884 \cdot 10^3 \sqrt{L \text{ (henry)} C \text{ (farad)}} \\ &= 59.57 \cdot 10^3 \sqrt{L \text{ (millihenry)} C \text{ (microfarad)}} \\ &= 1.884 \cdot 10^3 \sqrt{L \text{ (microhenry)} C \text{ (microfarad)}} \\ &= 1.884 \sqrt{L \text{ (microhenry)} C \text{ (micromicrofarad)}} \\ &= 1.884 \sqrt{L \text{ (microhenry)} C \text{ (picofarad)}} \end{aligned}$$

In connection with frequency and wave length of radio waves, it is interesting and helpful to keep in mind that

Velocity = Frequency (kilocycles) times Wave Length (meters) = 30,000 kilometers per second.

Velocity = Frequency (megacycles) times Wave Length (meters) = 300 megameters per second.

By these above simple numerical relations between frequency and wave length, one can easily find wave length if one knows frequency. If, for example, frequency is 1500 kilocycles (a radio broadcast frequency), the wave-length is 200 meters. ($200 \times 1500 = 30,000$).

A particular good and simple relation to be kept in mind is that 17 meters correspond to 17 megacycles. To be more accurate, the figures should be 17.32 meters and 17.32 megacycles but the 17 and 17 combination is an easy and good one to remember. If one keeps the number 300 in mind, then 1 meter quickly means 300 megacycles and 1 megacycle quickly means 300 meters. From this figure of 300, combinations of megacycles and meters are quickly calculated.

CHAPTER 11

Exponential Equations and Curves for Transient Conditions in Resistance, Inductance and Capacitance Circuits

The basic doctrine and natural law of geometric retrogression and progression are most interestingly illustrated in the increase and decrease of current and voltage in certain electrical circuits. When a constant voltage is connected and disconnected in inductance-resistance and capacitance-resistance circuits, transient or changing values of current and so called counter or "back" voltages result. The independent variable in these cases is time in seconds, milliseconds or microseconds. The changing state of affairs from the instant the circuit is closed ($t = 0$) until a final steady state of affairs is reached ($t = T$) is called a transient state. A transient state also prevails when the constant voltage is removed and the circuit is open. The curves portraying the rise and attenuation are called transient curves.

A SIMPLE APPROACH TO THE TRANSIENT CURVES

The problem of attenuation is well illustrated in Chart 3, page 188. The equation $i = Ie^{-\frac{Rt}{L}}$ will be recognized at once as typical of geometric retrogression. It need be looked at with no misgivings or fright. The "e" and the negative exponent are not horrendous. As the reader knows by now, the "e" has no mystical import.

As was done many times in the first chapters of this book, we may readily put the equation into the following form

$$i = I \left(\frac{1}{e} \right)^{\frac{Rt}{L}}$$

The $\frac{1}{e}$ is a sheer number; it is a fraction whose value is less than unity, namely .368. This gives the equation in the form $i = I.368^{\frac{Rt}{L}}$.

The .368 is the constant ratio between successive values of the dependent variable i , as time in its exponent increases.

If the time axis is given in numbers of the so called time constants $\left(T = \frac{L}{R} \right)$, then the equation becomes $i = I.368^T$.

If t in the exponent $\frac{Rt}{L}$ is assumed $\frac{L}{R}$ then $\frac{R}{L} \times \frac{L}{R} = 1$. The exponent of .368 is 1, and .368¹ is the fraction of I after a time of $\frac{L}{R}$.

After one time constant ($T = 1$), $i = .368 I$. After 2 time constants ($T = 2$) $i = I.368^2$. Now .368 times .368 = .135. When $T = 3$, then $i = I.368^3$. Now .368 $\times$.135 = .368 times .368 times .368 = .0496. Also .368⁴ = .018 and .368⁵ = .007. So i has the following values as a function of the time constant T , .368 I , .135 I , .0496 I , .018 I and .007 I .

After getting this simple approach and having a rational understanding of the whole situation, one very sensibly gets a bit of faith. He then asks logarithms to help out and with a sense of mastery not mystery use the equations

$$i = I \cdot e^{-\frac{Rt}{L}} = I \cdot 10^{-.4343 \frac{Rt}{L}}$$

$$e = 2.7183 = 10^{.4343}$$

$$\frac{1}{e} = \frac{1}{2.7183} = \frac{1}{10^{.4343}} = 10^{-.4343} = .368$$

These latter equations merely give directions as to how to get "i" by use of logarithms rather than a lot of multiplication.

In both cases of attenuation and growth, the exponent is really positive. However, as a matter of mathematical notation and machinery the negative sign may appear in the exponent to specifically denote attenuation.

In Chart 1 (Fig. 63), the equation is given as

$$i = I \left(1 - e^{-\frac{Rt}{L}} \right) = I \left(1 - 10^{-.4343 \frac{Rt}{L}} \right) = i = I \left(1 - .368^{\frac{Rt}{L}} \right).$$

The second term in the parenthesis is a factor that is also really attenuating in geometric retrogression fashion. As time progresses, the numerical value of this term gets small and smaller so that the numerical value of the entire expression in the parenthesis $1 - .368^{\frac{Rt}{L}}$ gets bigger and bigger. It is interesting to observe that this expression does not however increase in geometric progression fashion as time increases in steps of t . The various ratios (noted from curve in Chart 1) $\frac{.632}{0}$, $\frac{.865}{.632}$, $\frac{.950}{.865}$, $\frac{.982}{.950}$, $\frac{.993}{.982}$, $\frac{1}{.993}$, and $\frac{1}{1}$ are not equal and constant in value. The ratios decrease, and are infinity, 1.37, 1.09, 1.03, 1.01, 1.001 and 1.

EQUATIONS OF CURVES

Each curve in all the charts (1-17) has an equation, written alongside in most cases. Each equation expresses the relation of current, voltage, power, etc., as the case may be, at all instants of time from the beginning when switch is closed through the transient state of affairs to the final state of affairs. Some of the curves portray the transients after the switch is opened. The one common factor in the equations of all the curves is time (t) in seconds. Again it is to be said, that each equation gives the value of current, voltage, etc., at each instant of time as time goes on to the steady state of affairs when $t = T$. After a period of elapsed time (T), usually amounting to fractions of seconds when inductance or capacitance is very, very large, the transient state of affairs is over and a steady state of affairs is then reached.

All the equations are written in three forms as discussed on previous page.

The electrical phenomena involved in the simple circuits of resistance, inductance and capacitance obey the "natural" law of geometric retrogression. The curves of Charts 1-17 in later pages portray the principles discussed in the text.

With the confidence and assurance achieved by study and attendant understanding of geometric retrogression and exponentials, one needs to be unafraid and undaunted in further explorations. One appropriately asks at this stage: How does one arrive at the exponential equations; what basic and simple doctrines provide the approach; what fundamental truths and facts of electrical phenomena are invoked?

The answer is that, considering the simple facts offered by basic ideas, simple fundamental principles and actual experiments which make sense to one's understanding and into which one can get an insight by seeing and thinking about the modus operandi of experimental demonstrations, one can then do some more thinking and get more understanding about the problems in terms of symbols and mathematics. In other words, one makes use of some mathematics as a good tool and sees what happens. If one knows something about something in general and implicitly and then invokes formal mathematics to resolve his knowledge into a useful formula, then he knows more explicitly. The processes of mathematics enable one to arrive at conclusions which could not otherwise be reached. The engineer and skilled craftsman may need to ask the mathematician for help. To give help is what mathematicians are for. Mathematics cannot do anything for you unless you know something about something to start with and then ask mathematics (always a good tool) as such to help you out

and arrive at an explicit answer which you would not otherwise be able to get by means of just so-called common sense. This is the case in many situations of engineering and communication. Mathematics itself is of course developed by good common sense.

KIRCHHOFF'S LAW PROVIDES A SIMPLE AND BASIC APPROACH TO TRANSIENT EQUATIONS

Ohm's Law and Joule's Law express basic and simple truths about electrical circuits. The third in the triumvirate of principles relating to circuits which explicitly express basic truths are Kirchhoff's Laws. One of these latter two laws provides the approach to the question above stated, "Where do the exponential equations come from and how are they obtained."

The Kirchhoff Law or rather principle to be invoked states that the applied voltage in an electrical circuit is equal to the sum of the voltages aroused in the various parts of the circuit at all instants of time. If one had a simple circuit consisting of three resistances, the law says

Applied voltage = $E = R_1 I + R_2 I + R_3 I = \text{volts} = \text{ohms times amperes.}$

The $R_1 I$, $R_2 I$ and $R_3 I$ are called voltage drops.

INDUCTANCE

If one has a circuit of resistance and inductance, then E will equal the Ri drop plus the so called "back" voltage due to inductance. This latter voltage exists only when the current is changing; increasing or decreasing. The magnitude of this voltage depends upon product of the inductance L in henries and the time rate of change of current in amperes per second. The above two basic truths relating to the "backness" or direction of the self-inductive voltage and its magnitude are known as Lenz's Law and Faraday's Law. In terms of symbols and the code language of differential calculus, the time rate of change of current in amperes per second is conventionally written as $\frac{di}{dt}$. The di means a wee bit of current (i) and the dt means a wee bit of time, t .

If L were one henry and $\frac{di}{dt}$ were the instantaneous value of one ampere per second, at some particular instant of time, the voltage "aroused" would be 1 volt. At some other instant of time the $\frac{di}{dt}$ might be 2 amperes per second and the inductive voltage would be 2 volts.

When a simple circuit of resistance and inductance is closed with a voltage E applied, Kirchhoff's second law, expressed in terms of symbols, says $E = Ri + L \frac{di}{dt}$

The current (i) at $t = 0$ starts at zero (0) amperes and rises after some interval of time (T) to I . When the current rises to its maximum I , then there is no more $\frac{di}{dt}$. Then $E = RI$.

All the time from $t = 0$ to $t = T$, the current is changing. Its time rate of change $\frac{di}{dt}$ is very large and is a maximum at the beginning when $t = 0$ and decreases to zero (0) when $t = T$. At time ($t = 0$), there is no current (i) but a lot of $\frac{di}{dt}$ and $E = L \frac{di}{dt}$ since Ri equals zero (0).

So one can replace the E , which is constant, by RI and write the previous equation as follows:

$$RI = Ri + L \frac{di}{dt}$$

Therefore

$$I = i + \frac{L}{R} \frac{di}{dt}$$

$$i = I - \frac{L}{R} \frac{di}{dt}$$

This latter equation is appropriately called a differential equation because it expresses a basic truth about the differences and relations between i and $\frac{di}{dt}$.

When $\frac{di}{dt}$ (amperes per second) is very large as it is at beginning of time, i is zero. When $\frac{di}{dt}$ is zero at the end of the time (T) when the current does not change, i has risen to I .

At this point in the analysis, formal mathematical processes offer their talent and come up with the solution of the equation and provides the final equation.

$$i = I \left(1 - e^{-\frac{Rt}{L}} \right) = I \left(1 - \frac{1}{e^{\frac{Rt}{L}}} \right) = I \left(1 - .368^{\frac{Rt}{L}} \right) = I - I.368^{\frac{Rt}{L}} = I - I10^{-.4343 \frac{Rt}{L}}$$

The above simple and basic differential equation says that the current (i) in the circuit at every instant of time during the transient period including the initial instant of zero time, is equal to the final current (I), minus "some amount." The "some amount" at $t = 0$ is of course the I itself and $I - I = 0$ and so $i = 0$. It is likely, since a natural phenomenon is involved, that the "some amount" to be subtracted will be large (I) at zero time and gradually decrease or attenuate in magnitude to zero (0) at $t = T$, according to the doctrine of geometric retrogression. The formal mathematical solution of the differential equation substantiates one's prediction.

By writing the solution $i = I - I.368^{\frac{Rt}{L}}$, the above point is brought out by inspection.

If t were $10 \frac{L}{R} = T$, then $I.368^{10} = .000045 I$ and i will equal I . As shown in Chart 1, i is .993 I after only $5L/R$.

VOLTAGES

Of course, the voltages Ri and $L \frac{di}{dt}$ whose sum is always equal to E are always changing. At $t = 0$, when $i = 0$, Ri or e_R is zero and rises to a maximum of RI at $t = T$. The $\frac{di}{dt}$ (back voltage) is equal to E at $t = 0$ and becomes zero (0) when $t = T$. At $t = 0$, i is zero but it is changing (increasing) at its maximum rate. The $\frac{di}{dt}$ is maximum. At $t = T$, $i = I$ but has no time rate of change. The time varying voltages e_R and e_L are portrayed in Chart 2. The e 's and E are voltages in this case; the e is not the logarithmic base.

CAPACITANCE

In the case of the simple circuit of a battery with voltage, E and a resistance (R) and a capacitance (C), Kirchoff's principle also provides the basic approach. Applied $E =$ voltage drop of resistance plus the voltage acquired by the capacitor. The capacitor acquires a back voltage (e_c) as it were because it has received a charge (q) due to the current flowing into it. This voltage (e_c) at a particular instant of time is $\frac{q}{C}$. If the acquired charge (q) were one coulomb at this instant of time and the capacitance one farad, the voltage it has acquired would be one volt.

The above basic and simple fact can be stated in symbols

$$E = Ri + \frac{q}{C}$$

The instantaneous current i in amperes is now to be expressed in its equivalent coulombs per second $\left(\frac{q}{t} \right)$. This i becomes $\frac{dq}{dt}$ where dq means a little bit of charge (q) and dt means a bit of time, t . Again $\frac{dq}{dt}$ is an instantaneous value expressed in coulombs per second or amperes per second per second.

Therefore

$$E = \frac{q}{C} + R \frac{dq}{dt}$$

At the end when $t = T$, the charge on the capacitor is its final charge and is equal to Q , and its acquired voltage is E . Then $E = \frac{Q}{C}$ because current is zero (0) and therefore Ri is 0. The above equation becomes

$$\frac{Q}{C} = \frac{q}{C} + RC \frac{dq}{dt} \text{ and } q = Q - RC \frac{dq}{dt},$$

a differential equation.

At this point again in the analysis, formal mathematical processes solve the differential equation and provide the final equation.

$$q = Q \left(1 - e^{-\frac{t}{RC}} \right) = Q \left(1 - \left(\frac{1}{e} \right)^{\frac{t}{RC}} \right) = Q \left(1 - .368^{\frac{t}{RC}} \right) = Q - Q.368^{\frac{t}{RC}}$$

As before in the case of the rising current in an inductive circuit so this differential equation also says that the charge (q) on the capacitor at every instant of time during the transient period, including the initial zero instant of time, is equal to final charge (Q) minus "some amount." The "some amount" at $t = 0$, is of course Q itself and $Q - Q = 0$, and therefore $q = 0$. Similar to the inductive case, the "some amount" to be subtracted at $t = T = 10RC$ is zero (.000045 Q). So $q = Q$.

Again it is likely, since a natural phenomenon is involved, that the "some amount" to be subtracted will be a maximum at zero time and gradually decrease or attenuate in magnitude to zero (0) according to the doctrine of geometric retrogression. The solution of the differential equation again substantiates the prediction.

By writing the solution, $q = Q - Q.368^{\frac{t}{RC}}$, the above point is brought out by inspection.

This final equation is portrayed in Chart 6, page 194. The sudden increase and then decrease of current (amperes) occurring during the charging of the capacitor is portrayed in Chart 5, page 192. It is a fine and typical illustration of geometric retrogression. The current is a maximum at $t = 0$ but the charge (q) on the capacitor is zero at $t = 0$. Of course, the voltage acquired by the capacitor is also zero at $t = 0$.

In the study and mastery of transients, simple and basic truths about electrical circuits have provided the facts for mathematics to work on. Arithmetic, algebra, calculus and differential equations have been asked to assist. Calculus and differential equations are not easy. They are a bit tough in their advanced stages but are neither too difficult nor too highbrow for ordinary mortals who will give them a good try. In their basic and first principles, they are readily comprehensible and beautifully provide not only a plausible and sensible "take off," but also the final equation.

There is one fact that is quite interesting about the epsilon (ϵ or e), in all the equations. It is constant and always has the numerical value of 2.7182818284 (2.72 for short). How this constant "gets into the act" comes from the techniques of mathematics as has been previously discussed, page 133. The same is true of the constant pi, 3.1415926 (3.14 for short). Pi (π) is the ratio of the circumference of a circle to the diameter. The origin of the numerical value of "epsilon" is readily understandable by ordinary mortals who are not "math" experts. Its origin and numerical value were discussed in previous pages.

Suppose one has $1 + \frac{1}{2}$ raised to the 2nd power, namely $(1 + \frac{1}{2})^2$ or $(1 + \frac{1}{2})$ squared. $(1 + \frac{1}{2})^2$, as the reader can check, equals 2.25. Now $(1 + \frac{1}{3})^3$ or $(1 + \frac{1}{3})$ cubed, as the reader can check, equals 2.37. Also $(1 + \frac{1}{4})^4$, i.e. $(1 + \frac{1}{4})$ raised to the 4th power, is, as the reader can check, equal to 2.44. Now the reader can try $(1 + \frac{1}{10})^{10}$ and $(1 + \frac{1}{100})^{100}$ and $(1 + \frac{1}{1000})^{1000}$. These latter would involve a lot of arithmetic but if one did do it the value would approach 2.7182818284. This calculation was discussed in previous pages 145-148, but merits repetition now. Finally, if one had $\left(1 + \frac{1}{n}\right)^n$ where n is equal to a million, or ten million and finally "infinity," an immense number, the answer would be 2.7182818284. "e" and π never will stop having decimal places if one wants to keep up the calculation. "e" and π are not like $\frac{5}{4}$ which equals 1.25 and there is no need to calculate further.

Why, "e", equaling $\left(1 + \frac{1}{n}\right)^n$, where n is infinity, enters the drama is because of the sheer techniques of differential equations as these equations apply to the problem of a battery, a resistance, capacitance and inductance circuit. The simple facts and basic principles applying thereto supply the differential equations to start with and then the turning of the splendid wheels of mathematics provide an explicit answer involving, e .

The e , whose numerical value is 2.71828 and which comes from $\left(1 + \frac{1}{n}\right)^n$ where n is infinity, is the base for so-called natural or Naperian logarithms. (Take another look at page 133 and 101). Logarithms in general are exponents. The numerical value of 10 is the base for the plain and ordinary, variety of logarithms or exponents.

The mathematics which provide the final explicit equations for the L - R and C - R circuits is beautiful. The simple and basic electrical principles which provide original thinking and approach for the mathematics to work on are also beautiful. The explicit answers and the curves one gets are very important in communication and give valuable information about electromagnetic waves and oscillating circuits in carrier radio systems and other types of electrical communication gear and systems.

MATHEMATICS AND CONVERGENT SERIES

In passing, it is of interest to again mention how mathematics (the so called higher branches) provide answers by simple arithmetical means.

The circumference of a circular is generally accepted to be the diameter of the circle multiplied by the pi. As previously stated on a preceding page (86), the numerical value of π was given as 3.1415926 (3.14 for short). To get this value, it would be nonsense to try to measure the diameter and circumference of an actual easily drawn circle and then divide to get the ratio, π . One hundred people would get 100 different answers out in the 3rd or 4th decimal place. All one hundred might come up with $3\frac{1}{4}$ or 3.14 but it is worth while to get another way and have great accuracy.

The techniques of mathematics called convergent series evolving from several great theorems including those of McLaurin and Taylor provide an easy and wonderfully accurate method to find the numerical value of π .

The convergent series philosophy says

$$\pi = 4 \left(\frac{1}{2} - \frac{1}{3 \cdot 2^3} + \frac{1}{5 \cdot 2^5} - \frac{1}{7 \cdot 2^7} + \frac{1}{3} - \frac{1}{3 \cdot 3^3} + \frac{1}{5 \cdot 3^5} - \frac{1}{7 \cdot 3^7} \right)$$

By using only 9 terms of the first group and only 5 terms of the second group

$$= 4 (0.46367 + 0.32174)$$

$$= 3.1415926$$

A good approximation for π comes from the following:

$$= 4 \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} \text{ etc.} \right)$$

By the same convergent series method epsilon (ϵ) can be easily and very accurately calculated without the laborious arithmetic previously suggested

$$\begin{aligned} \epsilon &= 1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} \\ &= 2.7182818284 \end{aligned}$$

Incidentally the logarithms of numbers found in tables of logarithms are found so accurately by the same convergent series technique.

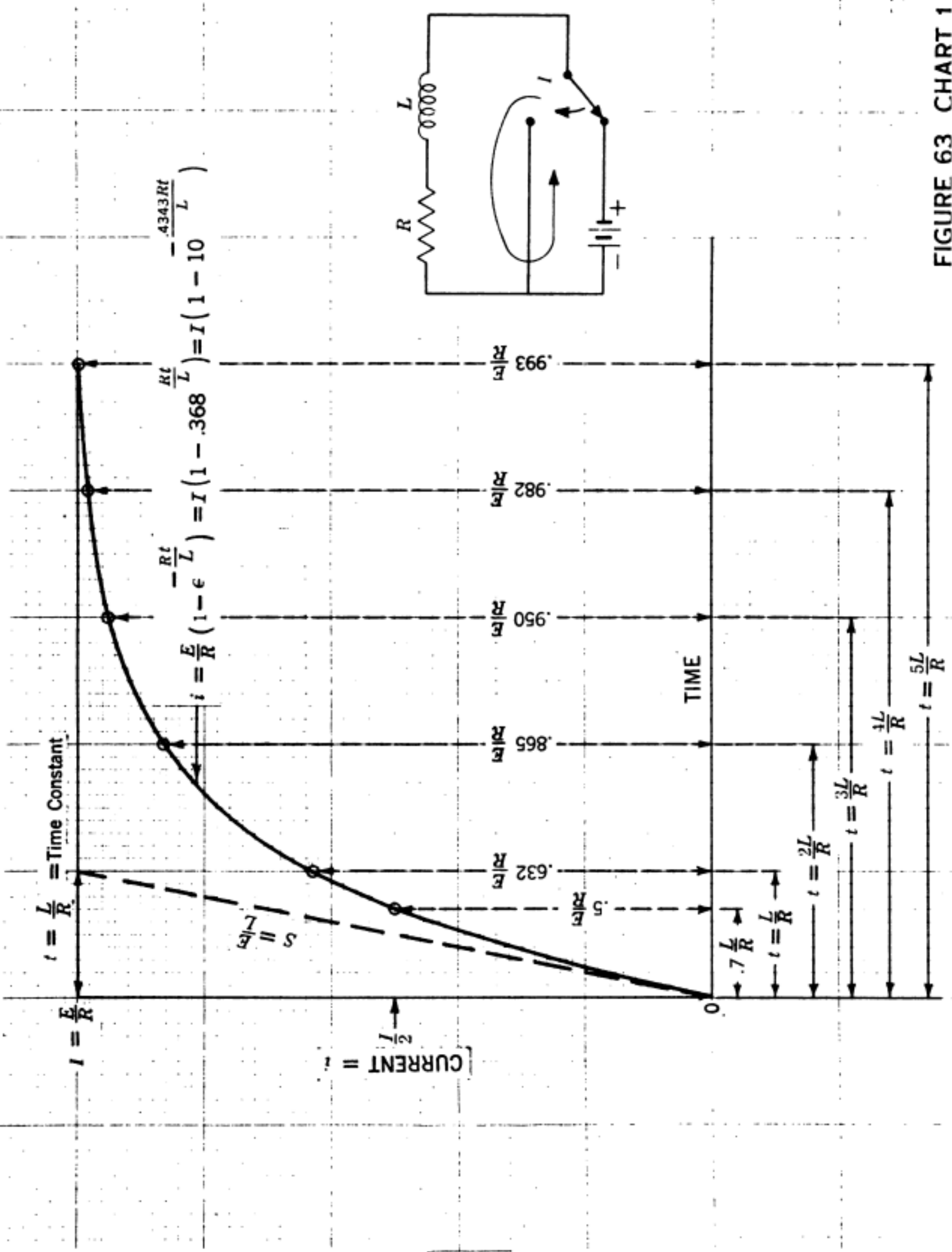
Some Observations about Transient Charts

CHART 1; Figure 63

The accompanying Fig. 64 sketch shows the circuit arrangements of E , L and R . When switch is closed, Chart 1 shows the rise or growth of current in the circuit. The curve starts out steeply, meaning the time rate

FIGURE 63 CHART 1

RISE OF CURRENT WHEN ENERGY IS STORED IN INDUCTANCE



RISE OF CURRENT WHEN ENERGY IS STORED IN INDUCTANCE

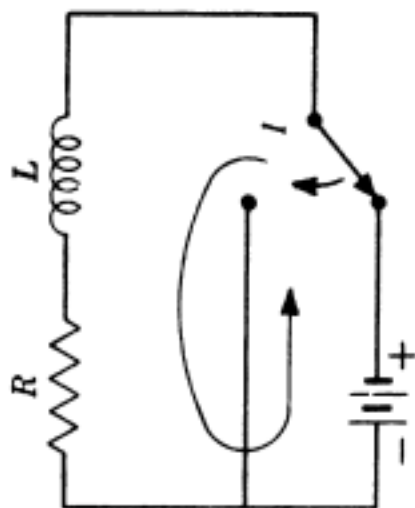
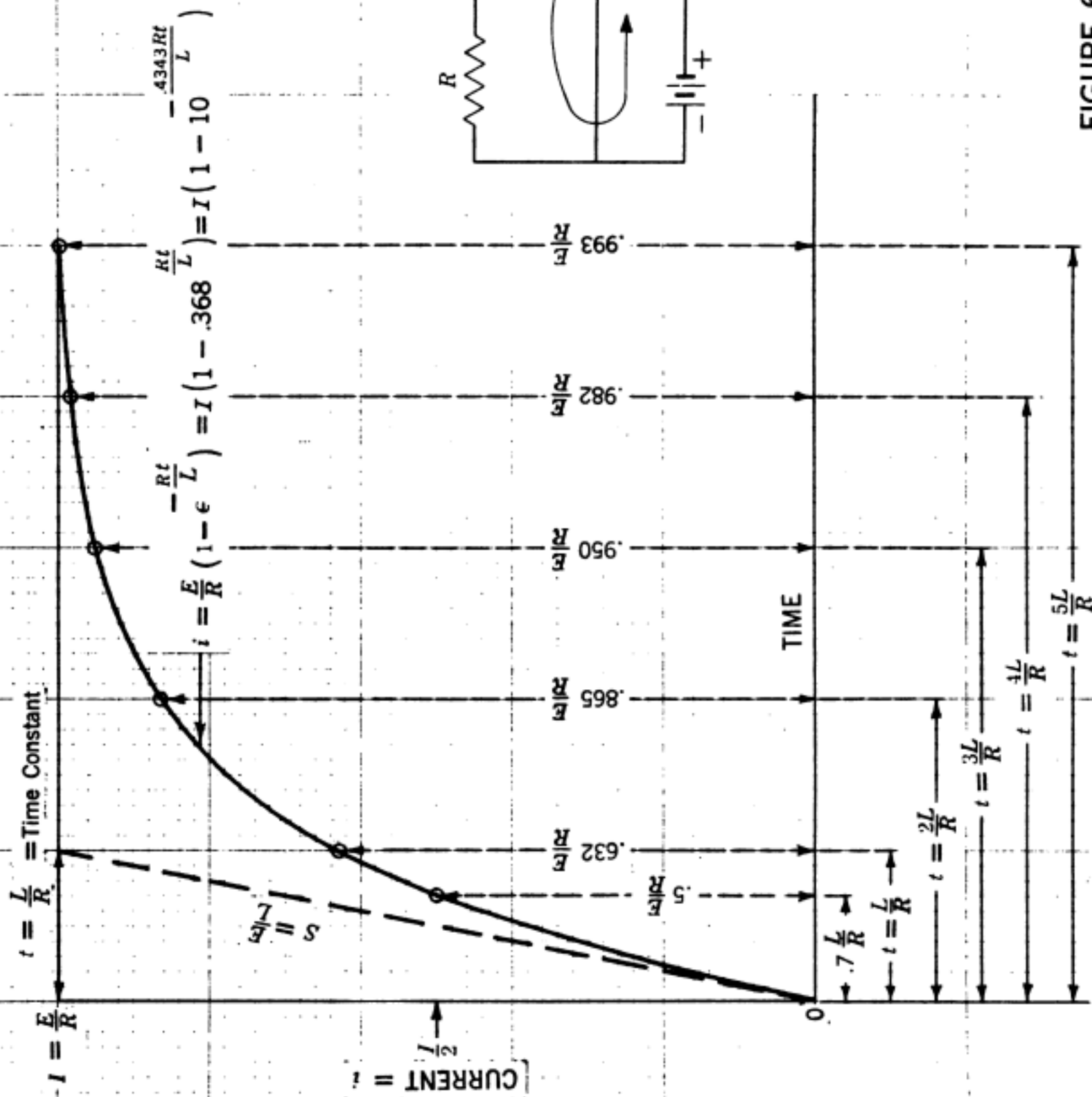


FIGURE 63 CHART 1

RISE OF CURRENT WHEN ENERGY IS STORED IN INDUCTANCE

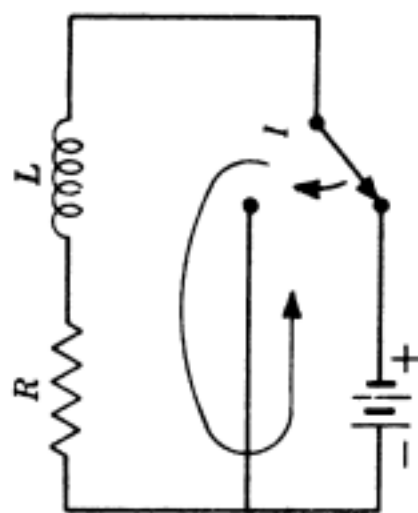
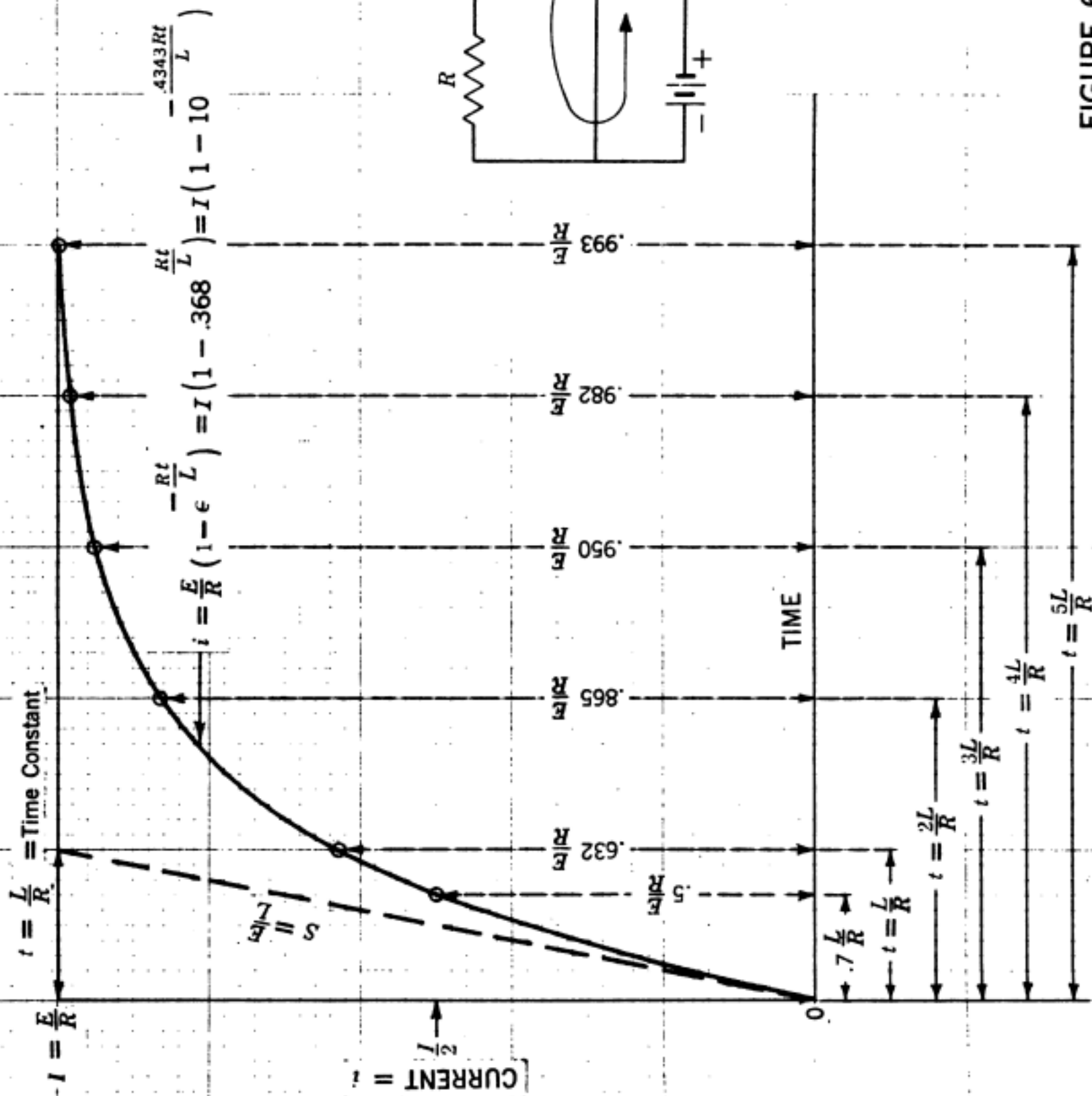


FIGURE 63 CHART 1

FIGURE 63 CHART 1

RISE OF CURRENT WHEN ENERGY IS STORED IN INDUCTANCE

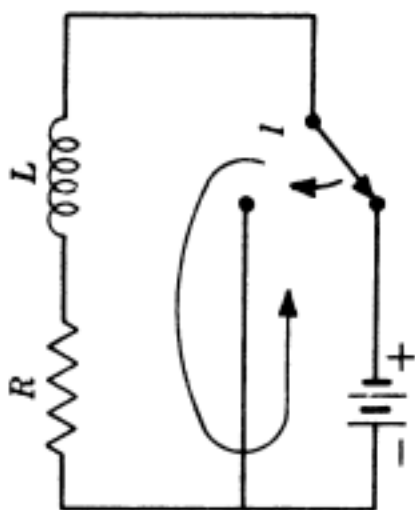
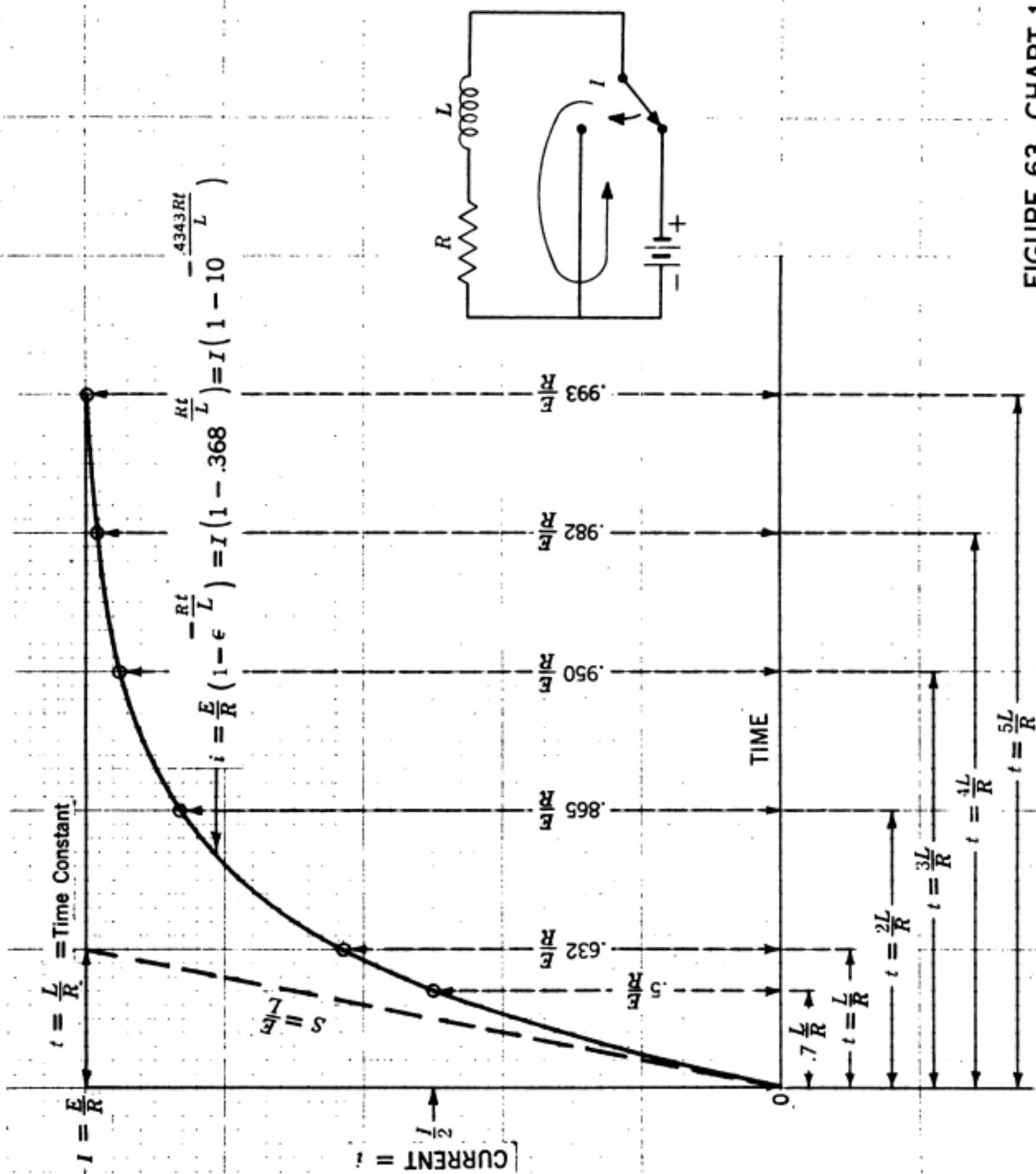
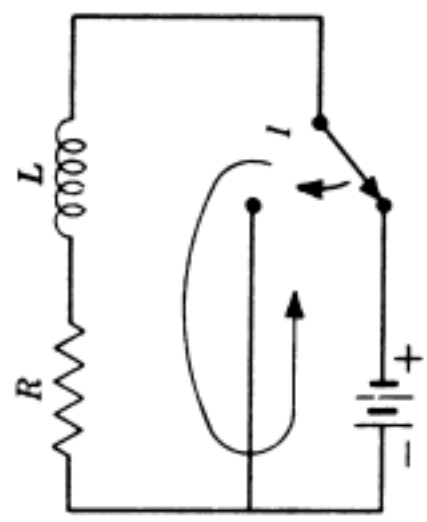
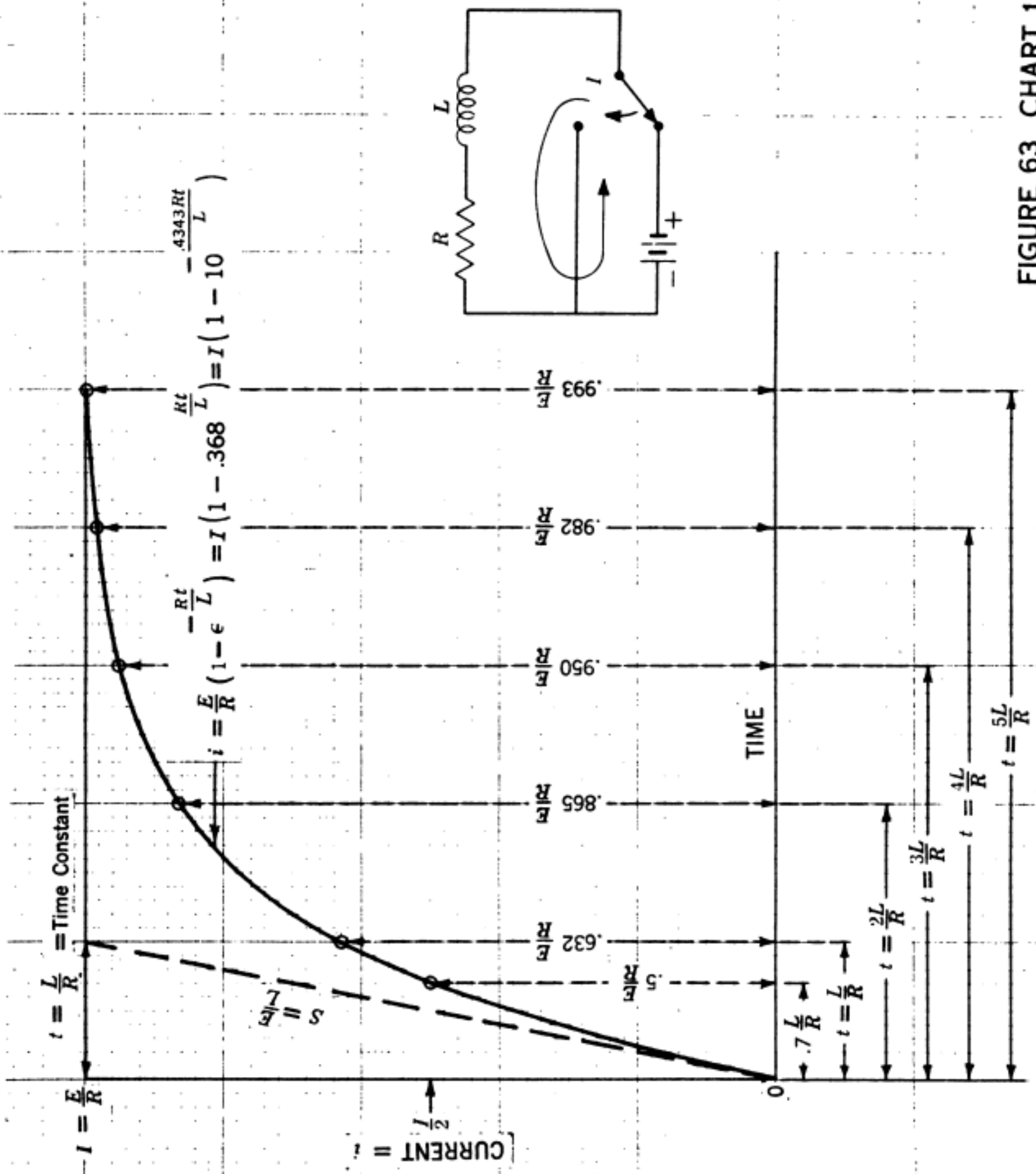


FIGURE 63 CHART 1

RISE OF CURRENT WHEN ENERGY IS STORED IN INDUCTANCE



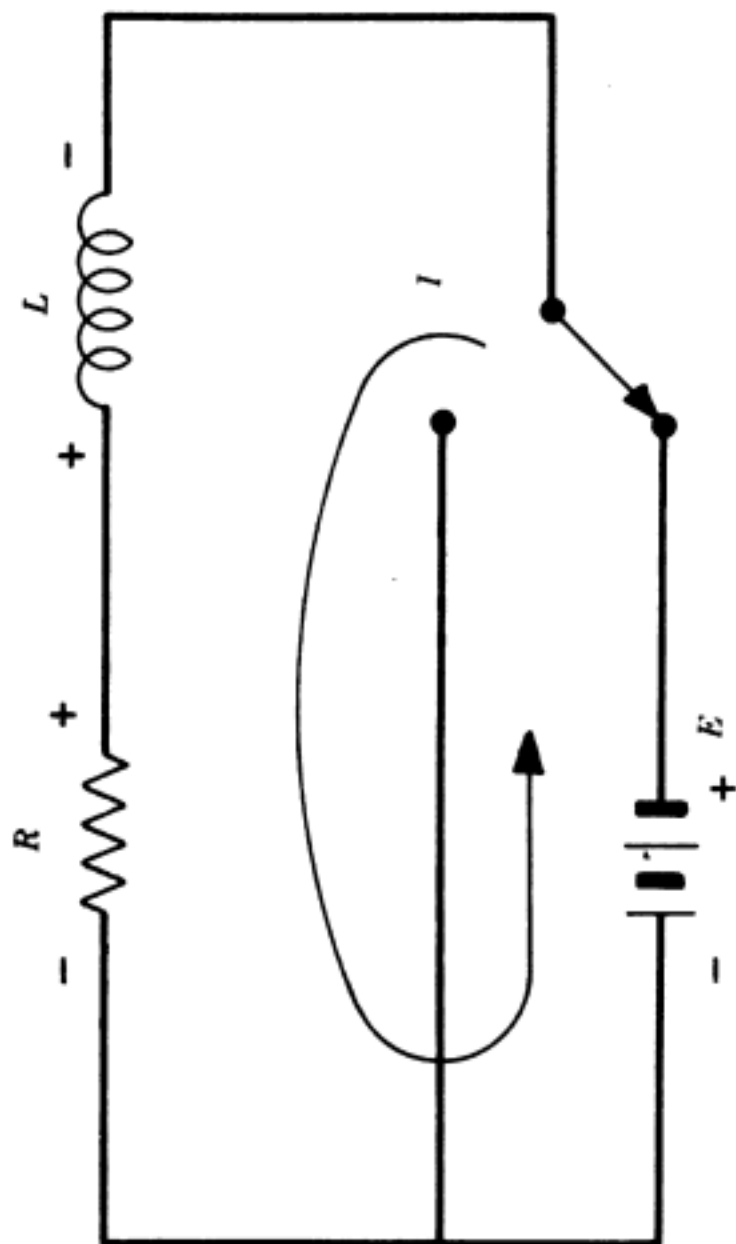


CHART 1 CIRCUIT
FIGURE 64

of change of current, the s , the amperes per second, is high. As time goes on, the curve becomes less steep. The time rate of change of current, the amperes per second (the s) decreases. The amperes as such however keep on increasing to some final steady state value equal to E/R or I , when t becomes T , namely $5\frac{L}{R}$ or better $10\frac{L}{R}$.

As indicated on the chart, varying values of elapsed time are shown on bottom axis of chart and expressed in $t = L/R, 2L/R$ etc. intervals. It will be seen that after an elapsed time of $5L/R$, the current has risen to 99.3 % of the final value, E/R (I). If elapsed time is taken equal to $10L/R$, then current becomes 99.995 % of E/R . This $10L/R$ is the elapsed time when current closely nears its positively final value, I . The I is called the steady state value or final value.

Time Constant

Oftentimes in discussions of the matter at hand, the value of elapsed time, L/R , is called the "time constant." It is the value of elapsed time when the rising current reaches 63.2 % of its final steady value, I . There is no particular reason for doing this. There is no particular significance to the 63.2 % figure. If " t " is made to equal L/R , then the exponent of e in the equation is -1 . But again, there is nothing of special value or significance in the -1 exponent.

$$1 - e^{-1} = 1 - \left(\frac{1}{e}\right)^1 = 1 - .386 = .632$$

The exponent could just as well be -5 or -10 as far as calculations are concerned; e^{-5} and e^{-10} are just as easy to look up in a mathematical table as e^{-1} . The L/R value for time (t) and the 63 % value for current (i) is an arbitrary matter. To be sure, the L/R value for time provides some quantitative notion of what is going on, which is of course always good. Quantitative values and relations are always important and significant in all branches of science and engineering.

The writer thinks it is more interesting, significant and important to know time elapsing until the current reaches its nearly final steady state value, when $t = T$. He would rather call $10 L/R$, the time constant when $i = 99.995\%$ of I or even call $5 L/R$ the time constant when $i = 99.3\%$ I .

Significance of Time Constant

It was stated in the previous paragraph that the 63 % I value for increasing current existent after an elapsed time in seconds equal to L in henries divided by R in ohms was of no particular importance, except as a definite quantitative figure. However the fact that the current-time aspect is expressible in a simple ratio, L/R or $5 L/R$ or $10 L/R$, does contribute to understanding and "know how" about transients, about amount of current, voltage, power, etc.

The time constant does give some idea on a quick appraisal as to how the current increases in some cases and decreases in others. If the inductance, L , is big and resistance R is small, the L/R says in general that the current rises more slowly on the average and takes a longer time to reach its final value than if L were small and R large. See Charts 1A, 1B, pages 203, 204. This truth about elapsed time was brought out in the previous discussion about the three wheelbarrows (page 140). It will be recalled that the velocity of the wheelbarrow was likened to current and that it took a longer time to get the massive lead-loaded barrow up to final velocity than it took for the empty barrow. The mass (M) of the barrow can be likened to inductance (L).

The determination of a certain period of elapsed time to be called "time constant" in terms of L and R as given by the simple formula, $t = L/R$, provides another good example of the value of a formula. The above simple formula puts in symbolic or shorthand form a fact which one may appreciate qualitatively and of which one may have a general insight. The formula provides and adds valuable explicit and quantitative information. Take a good look at curves of Chart 1A, page 203.

THE AMPERES PER SECOND; THE S WHEN $t = 0$.

As has been previously stated, the current increase curve of Chart 1 starts out very steeply. At the beginning, when $t = 0$, the s , the amperes per second is greatest (S), even though the current (i) at this beginning instant

is 0. The steepness or the slope of the curve at $t = 0$, is shown by the straight slanting dashed line. It will be observed that this dashed line hits the top line at $t = L/R$. The steepness or slope of this line is $I \div L/R$ or $E/R \div L/R$. In turn, $E/R \div L/R$ is equal to E/L .

The equation of this straight line is $i = \frac{E}{L}t$. When $t = 0$, $i = 0$; when $t = \frac{L}{R}$, $i = \frac{E}{R} = I$.

This slope, $\frac{E}{L}$, is an expression or formula for the amperes per second at the instant of closing switch when $t = 0$. Chart 1A (page 203) is to be discussed later, but right now turn forward to this chart and see the slant line showing that the amperes per second (at $t = 0$) is greatest when L is smallest and is smallest when L is greatest. The E (voltage) and R (resistance) for the three curves of Chart 1A are the same, but the L 's are different. In other words, the $\frac{E}{L}$'s or the initial amperes per second are different and each has a different maximum (S). In general, the amperes per second at any instant of time in this discussion will always be denoted by the letter S . The S would be expressed in amperes per second or coulombs per second per second.

It is deserving of particular mention that the amperes per second at $t = 0$, for an inductive resistance circuit, is independent of the resistance R . The amperes per second at $t = 0$ is dependent upon E and L only. The R has nothing to say about " s " when $t = 0$. R does have something to say about total time, T , however.

This is to be readily understood since, at $t = 0$, there is no current and the reactive voltage $\left(L \frac{di}{dt}\right)$ due to inductance is itself equal to the applied voltage (E). The reactive voltage has no assist from Ri since $i = 0$. The " s " is a maximum (S); the $\frac{di}{dt}$ is a maximum which in turn means that the slope or tangent to the curve at $t = 0$ is a maximum. The resistance R , whatever its value, has not yet had a chance to play a part because the current i at $t = 0$ is 0.

If the reader will now also look forward to Chart 1B (page 204), he will observe that the three curves, corresponding to three different R 's, have different time constants L/R , L/R_1 and L/R_2 ; and have the same amperes per second " s " ($s = S$) at $t = 0$. Since the L 's are the same, they have the same $\frac{E}{L}$. As per Chart 1B, a single straight slant line applies to all three curves. To be sure there are three different time constants; namely, L/R , L/R_1 and L/R_2 , but the same straight slant line ($E/L = S$) applies to all three curves. Again this means that the amperes per second " s " ($s = S$) are the same, even though the R 's are different. See points 1, 2, and 3 on the same straight slant line and reflect.

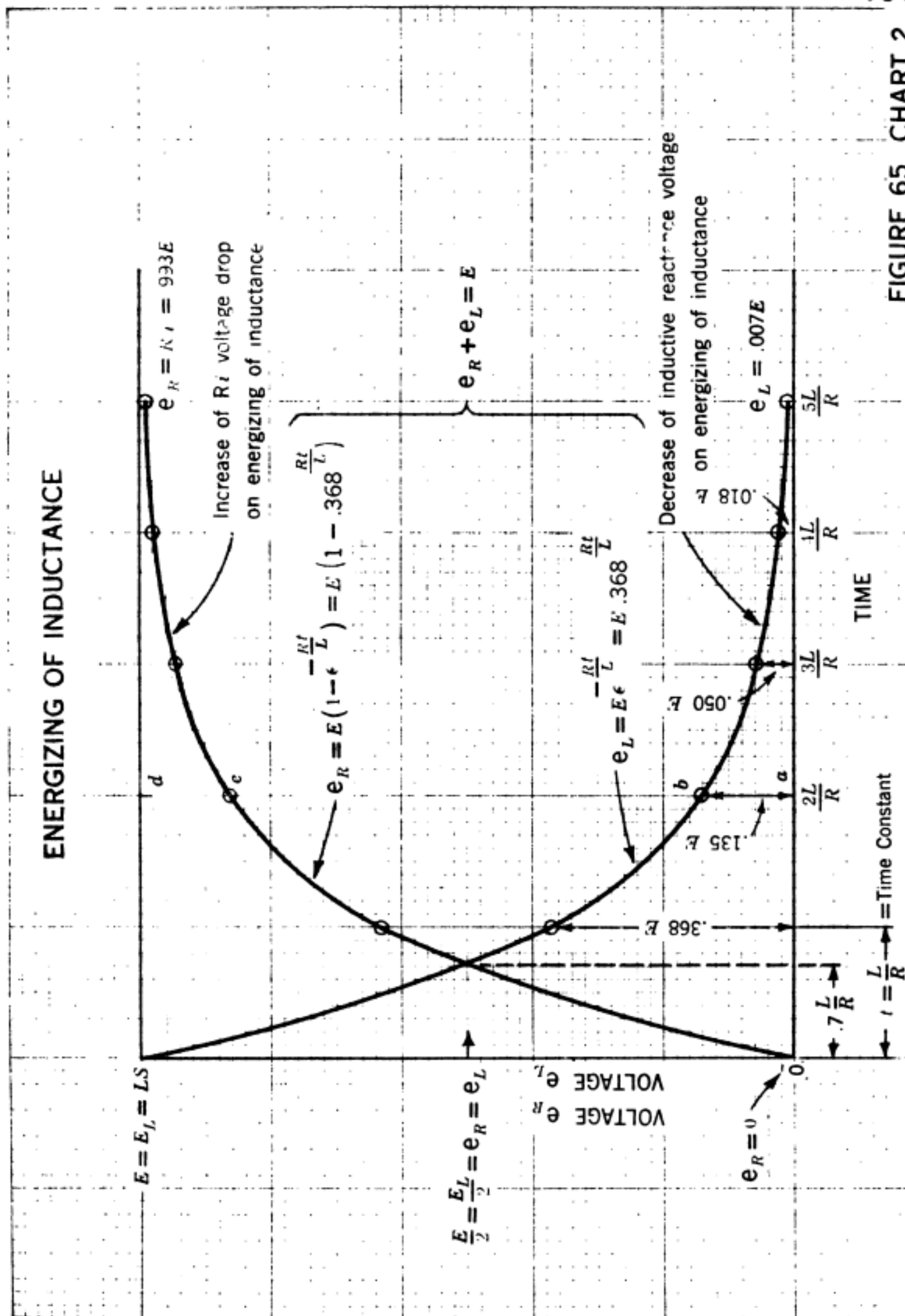
It has been previously stated that there was no particular significance to the fact that "time constant" is ordinarily defined as the time elapsing until the current rises to 63 % of its final value, I . It was pointed out that this definition was entirely an arbitrary definition. It was further pointed out that it might be better and more significant to define "time constant" as $5\frac{L}{R}$ or even $10\frac{L}{R}$. In $5\frac{L}{R}$ and $10\frac{L}{R}$ the current has very nearly reached the final current of $I = \frac{E}{R}$.

However, from the point of view of the straight slant line indicating the amperes per second at $t = 0$, the time constant L/R is a bit valuable as it can be used to give a good and new idea. One could validly say:

The "time constant," L/R , is the time it would take for the current to rise to its maximum, I , if it would continue to rise all the rest of the time at the same time rate with which it started to rise at $t = 0$. To put it another way, the time constant L/R may be defined as the time required for the current to reach the final maximum value if its initial time rate of increase (the amperes per second) were continuously maintained.

Again, the equation of the straight line is, $i = \frac{E}{L}t$. It is of course a hypothetical line.

The situation here reminds one of the wheelbarrow discussion earlier on page 140. If the pusher had a given amount of push that he would exert and if while giving the barrow increased speed, no wind or ground resistance were encountered, the velocity (current) would increase at a steady straight line rate up to a point of maximum velocity (current).



To be sure if there were no resistance in the electrical and mechanical case, the current and the velocity would continue to increase indefinitely. However, in both cases (inductance and wheelbarrows) the final current and velocity are controlled by a resistance so there is a limit to the increase in current and velocity respectively.

The "new" or at least different definition of time constant (L/R), as given above, nevertheless has a bit of interest.

CHART 2, FIGURE 65

Inspection of Chart 2 reveals some interesting and important facts about "back" voltage while energizing an inductance. These facts were qualitatively discussed and rationalized in the main part of this treatise.

At instant of closing switch when the current is zero, its time rate of change is greatest; the amperes per second ("s") is greatest, S (Chart 1). At this beginning instant ($t = 0$) the back or reactive voltage of the inductance and " S " is greatest and is actually equal to the battery applied voltage as Curve 2 shows.

The e_L , the reactive voltage, curve shows how the reactive voltage decreases from an initial maximum value as the current increases. To be sure, the current itself does increase but its time rate of increase ("s") decreases. The reactive voltage itself not only decreases but also decreases at a decreasing rate.

The voltage drop Ri or e_R is always directly proportional to i , the instantaneous current, so its curve has same shape as the curve 1 of Chart 1. The same kind of statement can be made about the changing e_r as was made about the changing current, i .

At all instants of time, the applied voltage of the battery E is equal to the sum of e_L and e_R . $e = e_L + e_R$.

An illustration of this point is to consider distance "ab" and distance "ac," on Chart 2, namely e_L and e_R at a certain instant. "ab" plus "ac" = "ad" = E . This fact is true for all instants of time as time increases along bottom "Time" axis.

At the point where the two curves cross, $e_L = e_R$ and of course here too $e_L + e_R = E$. Furthermore, since e_L and e_R are equal and their sum is E , it must be that e_R and e_L are each equal to $\frac{1}{2}E$ as observable on curves. The elapsed time when this is the case is $.7 L/R$ as shown on Chart 2. This latter fact about $.7 L/R$ as being the instant of time when $e_L = \frac{1}{2}E = \frac{1}{2}E_L = e_R$ is important. Since the current i at this same instant is also $\frac{1}{2}I$, the power delivered to the magnetic field as a reservoir at this instant is $\frac{1}{4}EI$ (25% EI). This EI is the power delivered by the battery at the last instant of steady state condition and continuously thereafter. This interesting fact of 25% EI is shown in Chart 11 and 12 (pages 210, 213) and is further discussed therewith.

Since e_R also equals $\frac{1}{2}E$ at $t = .7L/R$, the power delivered by battery to resistance (R) at this same instant ($\frac{1}{2}I$ times $\frac{1}{2}E$) is also $\frac{1}{4}EI$. This power is wasted. So the 25% power to L and the 25% power to R means that at $t = .7L/R$, battery is delivering 50% of its final power, EI . See Chart 12 (page 213) and note that p_R curve intersects p_L curve at the latter's peak.

Energy

Since the powers delivered to inductance and resistance at $.7 L/R$ are the same, the energies involved must be the same since the elapsed time is the same. This means that 25% of the energy has been wasted in heat and 25% of the energy is stored in inductance reservoir during the $.7 L/R$ seconds of elapsed time. See Charts 17, 17A Page 222.

So as stated above, 50% of the total energy involved up to final steady state is expended during the first small interval of time

It is also worth taking note that the p_r curve (Chart 12) page 213 is at $.5 EI$ at $t = .7 L/R$. The p_r ordinate distance at this point is sum of two equal distances, each representing $.25 EI$.

At elapsed time of L/R , the reactive, e_L , voltage has decreased to 63% of its initial maximum value and e_R has increased to 63% of its final value since the current (i) has also increased to 63%; 63% E/R .

At instant of closing switch, $E_L = E$ and $e_R = 0$. At final steady state (when $t = 10 L/R$) reactive voltage $e_L = 0$ and $E_R = RI = E$. i = final maximum value, I . The curves e_L and e_R show these facts.

In Chart 2, the e_L and e_R are therefore both drawn above the bottom base line. The constant E of the battery is represented at top of vertical reference line.

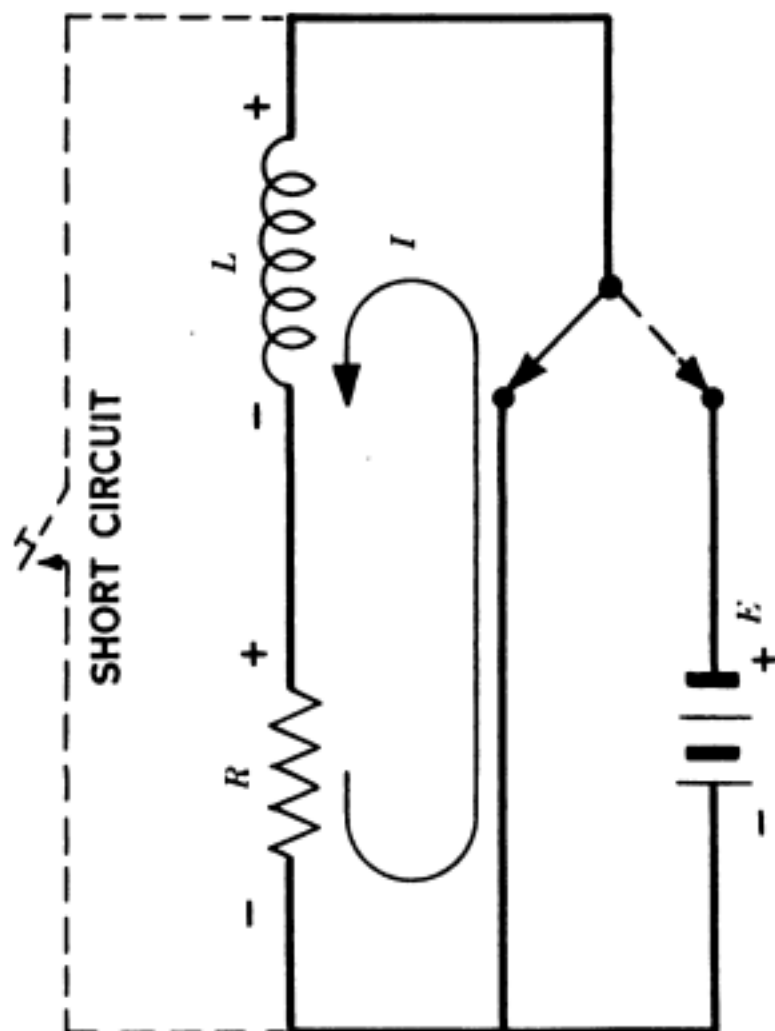


CHART 3 CIRCUIT
FIGURE 66

DECREASE OF CURRENT WHEN INDUCTANCE GIVES BACK ITS STORED ENERGY

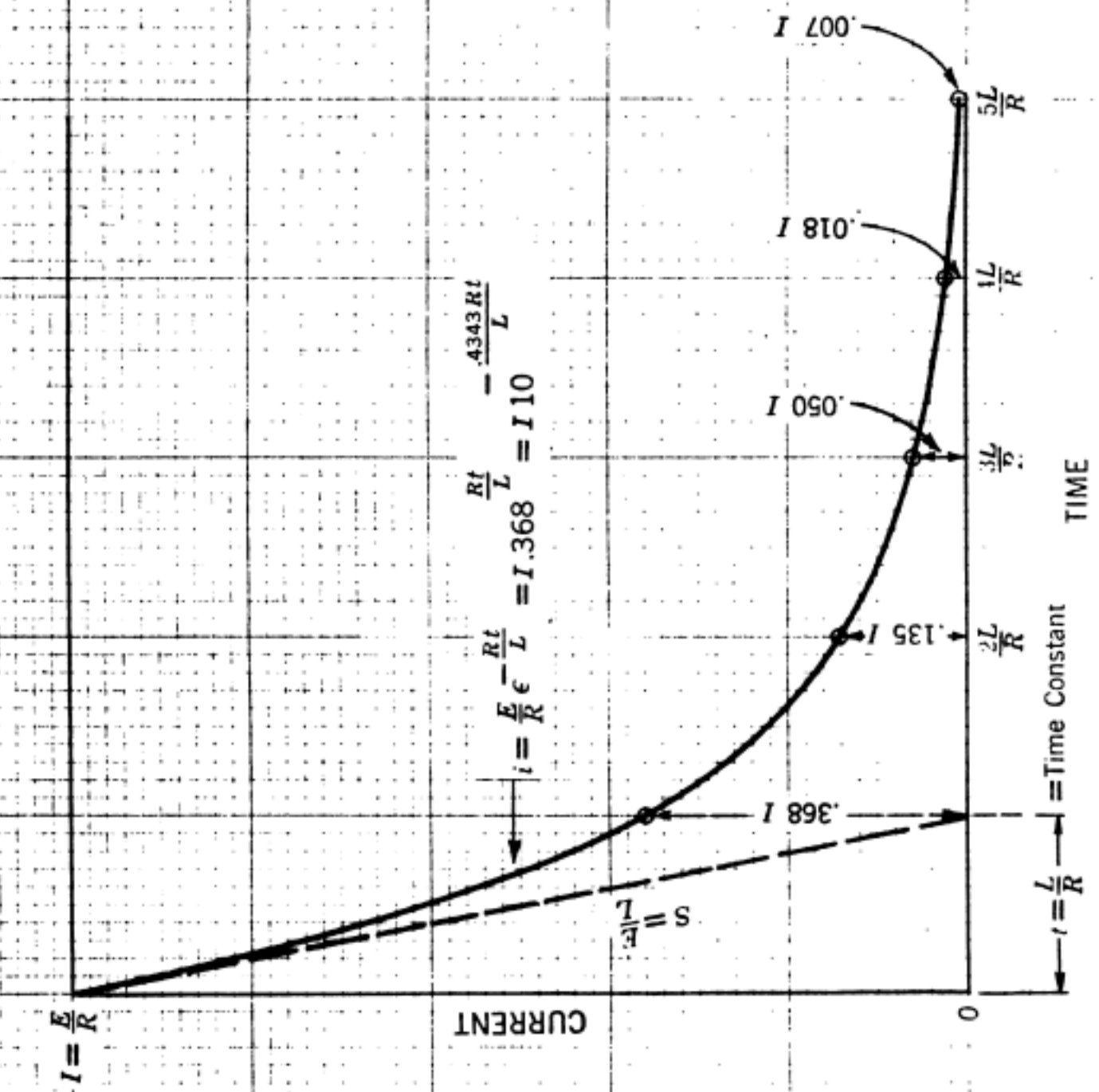


FIGURE 67 CHART 3

CHART 3, FIGURE 67

The sketch of circuit for Chart 3 indicates that the inductance-resistance battery circuit has been opened and then closed, without battery, to permit de-energizing of inductance. The things portrayed in good fashion by Charts 1, 1A, 1B, 1C (page 203) happen on closing of switch as shown in circuit sketch. These things actually do happen as a practical experimental fact. However, the events portrayed on Chart 3 are purely hypothetical, and impossible of attainment unless a short circuit is made on the whole system as indicated by dotted line in sketch while lower switch is kept closed in original charging position. If lower switch is opened to disconnect battery (short circuit switch open) and thrown upwards to short circuit the energized inductance, the question is what actually happens practically if the curve of Chart 3 is hypothetical as above stated. The thing that happens is, that at the very instant the original closed switch contacts are opened, there is a very high resistance in the new existing circuit before lower switch can be closed to upper contact. The switch in dotted line circuit is left open during these latter proceedings.

At the first tiny instant of time at opening of switch, the resistance in the new circuit is infinitely large. The lower switch in circuit sketch cannot be instantaneously opened and closed as in sketch to provide conditions portrayed in Chart 3. On initial closing of a circuit (a perfectly possible situation, as indicated in Charts 1, 2), the time constant was L/R where R is the resistance shown. But now at opening of switch, there is a new resistance R_0 , a very large R_0 in the circuit. Now the time constant is $\frac{L}{R + R_0}$. The new time constant on opening of switch is now very, very small because the resistance is very large. Time constant now has a new big denominator. The current, namely, the maximum current, I , (that has been flowing) would not decrease slowly as shown in Chart 3 but would drop to zero very quickly. If it drops to zero very quickly then the rate of change of decrease at the instant of opening, the amperes per second (S) at which it starts to decrease is not as shown in Chart 3, but is much greater. Its curve of decrease would be much steeper than shown. It does not actually take $5\frac{L}{R}$ to drop to practically zero as shown on Chart 3, but a very much shorter time. The importance of these facts is discussed in considering Chart 4. Chart 3 is a good chart in general but does not portray actual events as already mentioned and discussed more completely in next pages.

Short Circuit of Inductance Resistance

In stead of trying to open battery line switch and then closing it again as indicated in accompanying circuit sketch for Chart 3, one could put a short circuit (zero resistance) by closing upper key right across battery as indicated by dotted lines after steady state was reached. This would also, of course, short circuit the resistance-inductance. In this event, the R would not be changed and the current I flowing would not encounter any different resistance than that which has controlled it in the steady state, namely $I = \frac{E}{R}$.

Under this drastic procedure, the Curves 3 and 4 would portray the actual happenings regarding decrease in current and reactive voltage. The reactive voltage at instant of short circuit would not be greater than E , the applied voltage. The energy stored in the inductance would finally be dissipated in heat in R .

CHART 4, FIGURE 68.

The two curves of Chart 4 are actually impossible and entirely hypothetical as was curve of Chart 3 under the conditions of ordinary de-energizing. A short circuit procedure would provide validity to the curves of Chart 4 as it did for Chart 3.

If, according to previous paragraphs, the current does not decrease as shown in Chart 3, but decreases at a much faster rate and takes much less than $5 L/R$ time to drop to zero, then the reactive voltage of inductor associated with the very rapid decrease (collapse) of current (giving up of energy) at instant of opening circuit (a very large amperes per second reduction), must be much higher than the lower curve Chart 4 for E indicates. The lower curve indicates that the induced reactive voltage at instant of opening is equal to the applied battery voltage E , but actually in most cases it is much greater. One would need to go way below the low point of the Chart 4 to get this increased and short lived e_L .

**$e_R = iR$ DROP AND INDUCTOR VOLTAGE
VERSUS TIME WHILE DE-ENERGIZING**

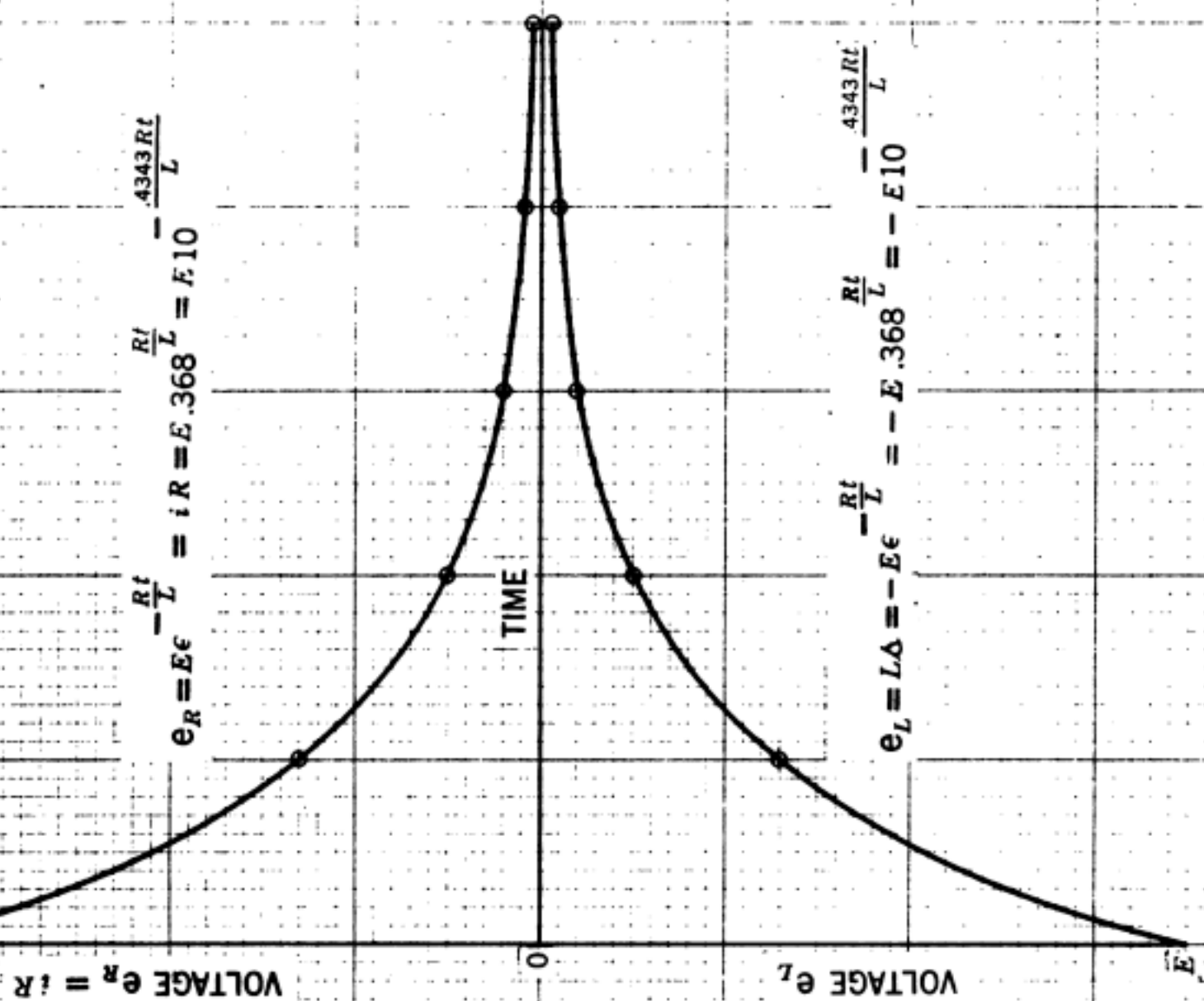


FIGURE 68 CHART 4

ENERGY VERSUS TIME IN AN RL CIRCUIT DURING ENERGIZING OF INDUCTOR

Time Constant = $\frac{L}{R} = .01$ Second

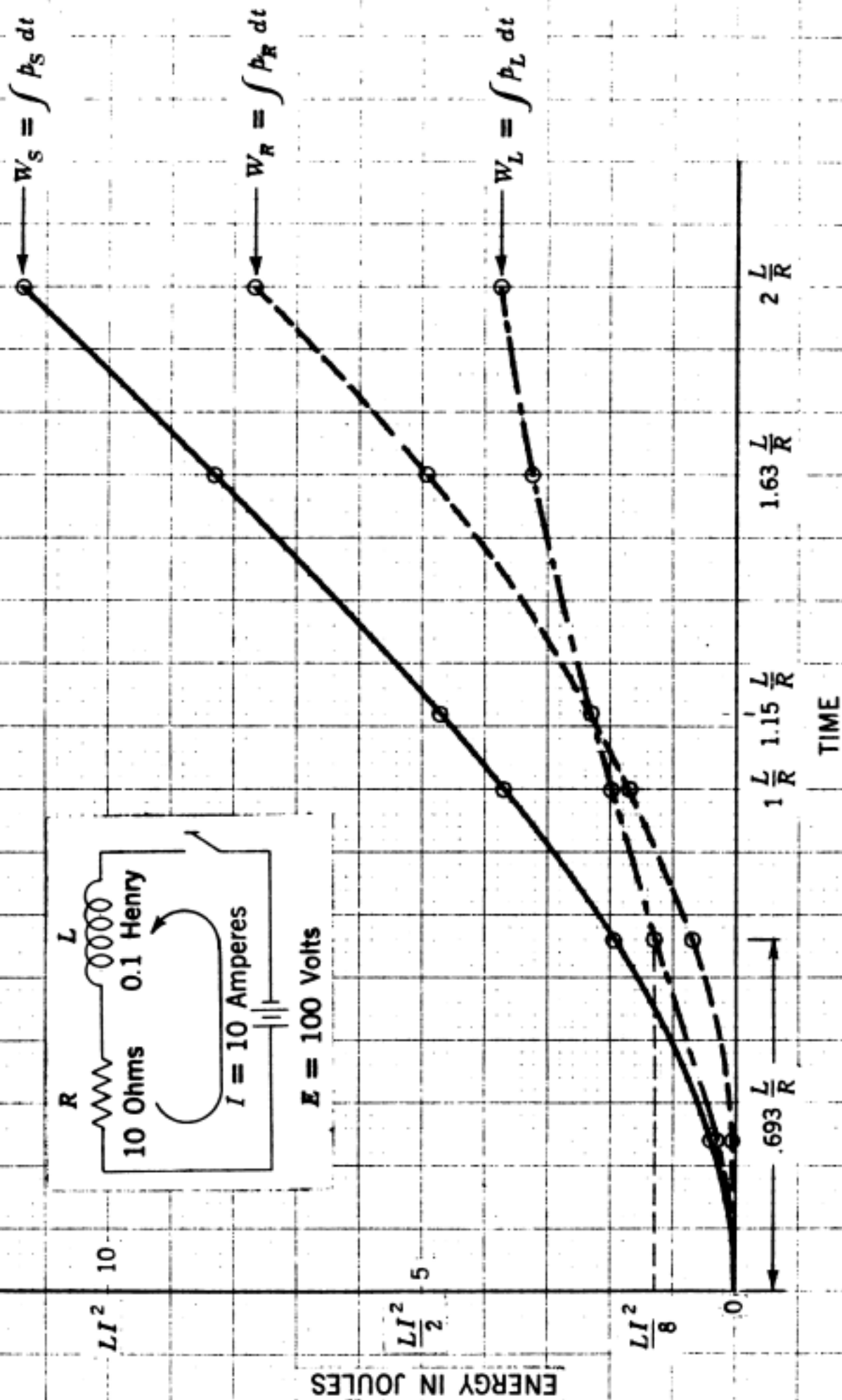
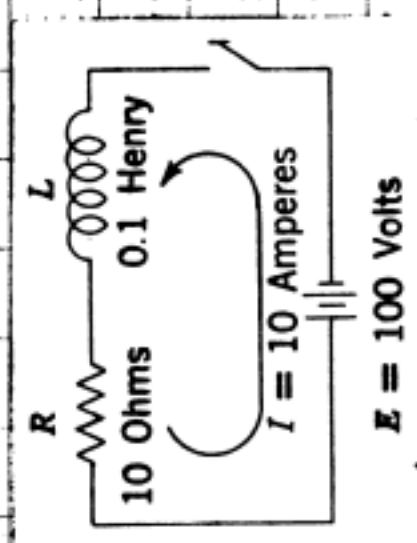


FIGURE 87 CHART 17

ENERGY VERSUS TIME IN AN RL CIRCUIT

CONTINUED FROM CHART 17
USING DIFFERENT SCALE

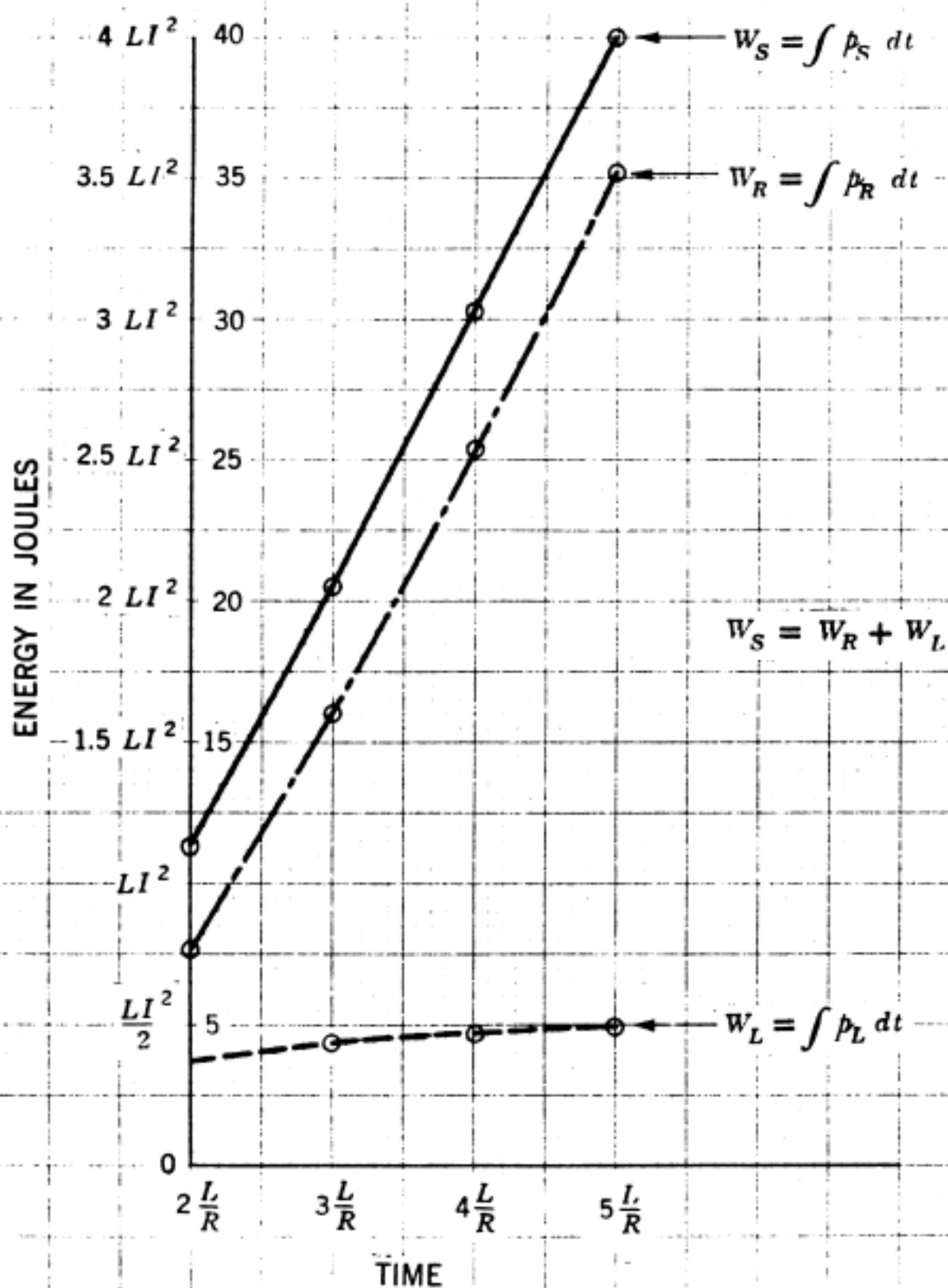
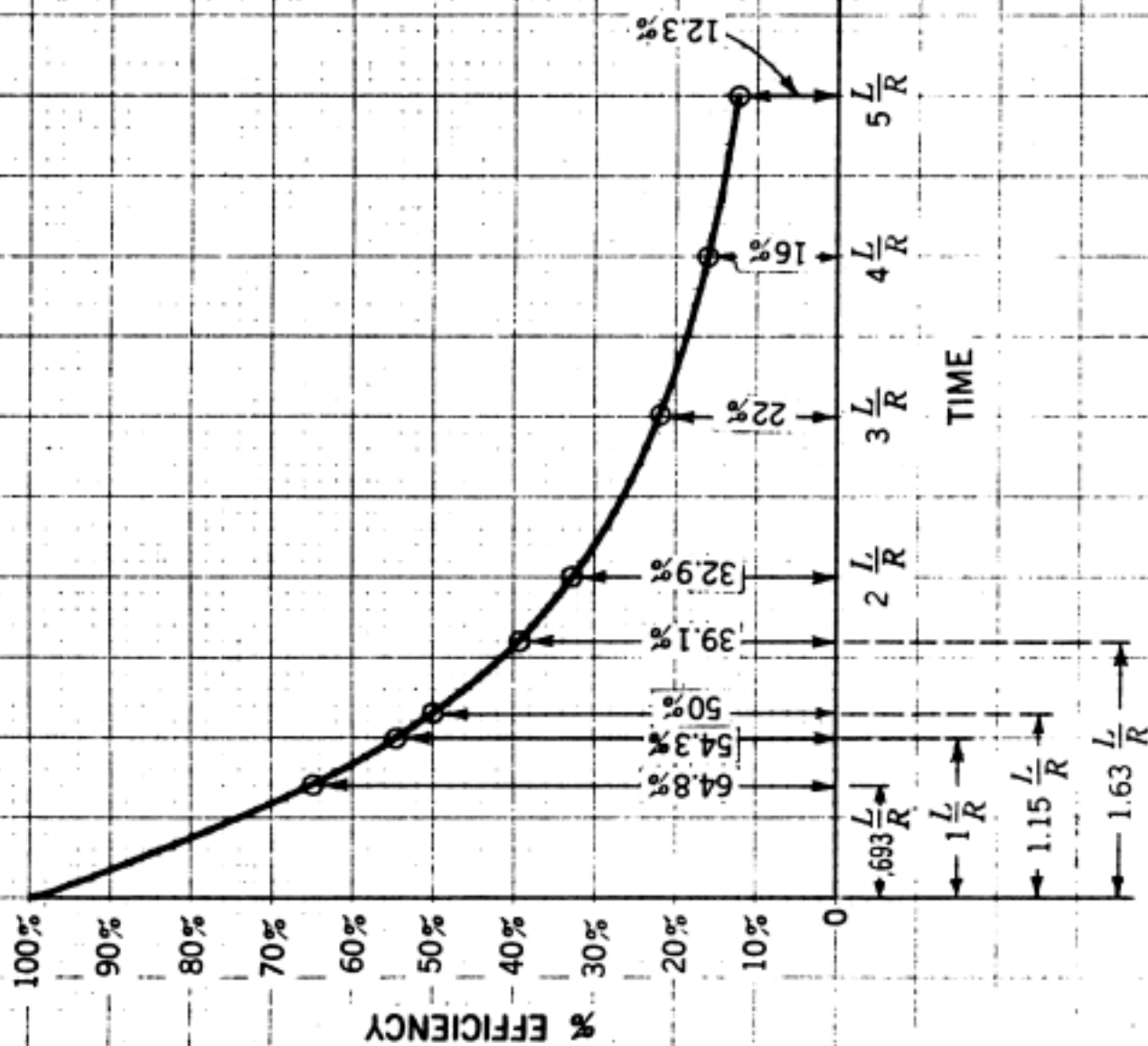


FIGURE 88 CHART 17A

FIGURE 89 CHART 18

EFFICIENCY VERSUS TIME IN AN RL CIRCUIT



CHAPTER 12

Exponentials, Vectors, Series Expansion for Sine, Cosine, Sinh and Cosh

The many contributions of the science of mathematics to the other sciences (physical, chemical, natural and biological) have been highly valuable and important. Mathematics provides explicit and quantitative information which enhances the understanding and usefulness of the qualitative character of natural phenomena.

In the earlier pages of this treatise, mathematics per se provided many "assists," as they say in baseball. These "assists" articulated basic and fundamental ideas and concepts relating to the four conic section curves in general and to the sine, cosine, sinh and cosh functions in particular.

The fine achievements of the series expansion doctrine have been invoked. This doctrine has made many excellent "assists". In this portion of the book, the mathematical doctrines of vectors and exponential expressions will be asked for more "assists". The principal objective is to get more understanding and more "know how" about a number of important factors which have and are contributing to the design, modus operandi and splendid development of apparatus and equipment in communication engineering. Many engineers and research specialists with their skill and competence comprise a great team which has won many victories in the progress of technological endeavor.

NUMERICAL VALUES

It has been previously pointed out that the actual and precise determination of the numerical values of sines and cosines for different points on the circle and the hyperbola with angles measured in circular radians and hyperbolic radians respectively is made possible by the use of series expansion technique. As stated before, the creative mathematical and development work of McLaurin, Taylor and others made the series expansion technique possible.

SINE, COSINE

In establishing and developing the series expansion expression or formula for sine and cosine, the specifications were laid down by the intrinsic nature and geometric relations of these two functions as provided and decreed in the simple equation of the circle.

$$\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1 = \cos^2 \theta + \sin^2 \theta$$

R is not only the radius of the circle but also is the hypotenuse of many triangles and plays the role of a constant reference distance for the formulation of the circular sines and cosines.

Whatever the series expansion formula for circular sine and circular cosine is, the numerical values forthcoming must satisfy or be roots as it were of the above equation for all possible combinations of X and Y as they exist and relate to various angles when expressed in circular radians. So thanks to the series expansion technique

$$\sin \theta \text{ or } \sin X = X - \frac{X^3}{3!} + \frac{X^5}{5!} - \frac{X^7}{7!} \pm$$

$$\cos X = 1 - \frac{X^2}{2!} + \frac{X^4}{4!} - \frac{X^6}{6!} \pm$$

These expansion expressions permit of easy and precise determination for numerical values of $\sin \theta$ and cosine θ where θ is in circular radians.

SINH, COSH, HYPERBOLIC FUNCTIONS

In establishing and developing the series expansion for $\sinh$ and $\cosh$, the specifications are laid down and decreed by the simple equation of the equilateral hyperbola $\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1 = \left(\frac{X}{a}\right)^2 - \left(\frac{Y}{b}\right)^2$

The nature or physical and geometrical relation of $\sinh$ and $\cosh$ are provided by $\frac{X}{a}$ and $\frac{Y}{b}$. See previous discussion about $\sinh$ and $\cosh$ as such, pages 66-77. Figures 35, 36, 37.

As developed in the discussion of the hyperbola, $\sinh$ and $\cosh$, the "a" is an "abscissa" and the "b" is an "ordinate". Actually they are equal in magnitude and are simply portrayed in the sketch of an equilateral or rectangular hyperbola on the X coordinate axis and in the conjugate equilateral or rectangular hyperbola on the Y coordinate axis respectively.

Each serves as a constant reference distance to which the X and the Y of points on the hyperbola can be compared. They serve in a capacity exactly analogous to the constant radius of the circle. (Constant hypotenuse of the triangle.)

If the "a" and the "b" are validly assumed to be equal to unity (1), then for the equilateral hyperbola, the equation becomes

$$X^2 - Y^2 = 1 = \cosh^2 \theta - \sinh^2 \theta.$$

Whatever the series expansion doctrine provides for $\sinh X$ and $\cosh X$, the numerical values forthcoming must satisfy or be roots of the above equation for all possible combinations of X and Y as they exist and relate to various hyperbolic radians, θ or X .

So thanks to the series expansion technique provided by skill of the mathematician

$$\sinh X = X + \frac{X^3}{3} + \frac{X^5}{5} + \frac{X^7}{7} + \dots$$

$$\cosh X = 1 + \frac{X^2}{2} + \frac{X^4}{4} + \frac{X^6}{6} + \dots$$

EXPONENTIALS, SINH, COSH

In the course of the thinking and the analysis about matters in general relating to sine, cosine, $\sinh$ and $\cosh$ and the actual precise evaluation of their magnitude, it occurred to the mathematical scholar that, instead of laboriously substituting various values of X in the many terms of the $\sinh$ and $\cosh$ series expansion to find their numerical value for different hyperbolic radians, he would save labor and effort by taking a look around and explore the series expansion for e , the Napierian logarithm base. This "e" bobs up in mathematics frequently and maybe it would offer its talent in the overall attack.

In general, "e" comes out of the binominal theorem, $e = \left(1 + \frac{1}{n}\right)^n$ where "n" is infinity. A great deal will be presented about "e" in later pages.

With the specifications decreed by "e" being $\left(1 + \frac{1}{n}\right)^n$, the series expansion technique comes up with the following

$$e^X = 1 + X + \frac{X^2}{2} + \frac{X^3}{3} + \frac{X^4}{4} + \frac{X^5}{5} + \dots$$

$$\frac{1}{e^X} = e^{-X} = 1 - X + \frac{X^2}{2} - \frac{X^3}{3} + \frac{X^4}{4} - \frac{X^5}{5} \pm \dots$$

If X is unity (1), $e^X = e = 2.718281828459$. $e^{-1} = .36787944$.

NUMERICS

The X in the above expression is a pure numeric; it is not something with quality or nature in terms of mass, length or time. It is merely an exponent of " e " and provides a numerical value for " e " to be raised to some power; square, cube, etc. (also a pure numeric). It cannot be a voltage, a linear velocity, an angular velocity, an acceleration which all have names, volts, miles per hour, radians per second or feet second per second respectively; the exponent X is a sheer numeric.

The " e " is a pure number and e^x is a pure number; they have no names. However the exponent X is sometimes called a neper.

One of the pure numbers oftentimes encountered is angle measured in degrees or radians. After all, an angle is a ratio; it is a numeric; it is the ratio between two distances which to be sure do have names. The arc along the circumference measured in inches (perhaps) is divided by the radius measured in inches too; the ratio is angle, a sheer number without a name. One might argue that the words radians and degrees are names but actually they merely express a ratio or comparison. They are like degrees of temperature. Temperature has no name; it is a numeric.

This fact about angle being a sheer number will be used later in this treatise.

Angular velocity (ω) does involve time and has therefore a name; radians per second. The angle θ itself or ωt has no name.

COMPARISON OF EXPANSION'S EXPRESSIONS

With a few appraising looks over the situation and comparing the e^x and e^{-x} series expansion and the sinh and cosh series expansions, one comes up by simple addition and subtraction of the expressions on previous pages with the following expressions

$$\sinh X = \frac{e^x - e^{-x}}{2} \quad \cosh X = \frac{e^x + e^{-x}}{2}$$

Writing these in a different form and doing some simple and straightforward rearranging

$$\begin{aligned} e^x &= \cosh X + \sinh X \\ &= \frac{e^x}{2} + \frac{e^{-x}}{2} + \frac{e^x}{2} - \frac{e^{-x}}{2} \\ e^{-x} &= \cosh X - \sinh X \\ &= \frac{e^x}{2} + \frac{e^{-x}}{2} - \frac{e^x}{2} + \frac{e^{-x}}{2} \end{aligned}$$

The upshot of the exploration is that the numerical values of e^x and e^{-x} obtainable from their expansion expressions, can be readily calculated once and for all. Then having these numerical values in a table, sinh and cosh in any required case can be readily calculated. Finally, after considerable arithmetic to be sure, a table for sinh and cosh can be made. Such tables are found in various texts and books of tables.

DEFINITIONS

In some treatises, one finds the statement that hyperbolic functions can be defined by the above expressions.

With this statement, the writer disagrees. The above equations do not define; they only provide a simple method of precise numerical evaluation. Sinh X and cosh X are however truly defined by considering their nature and geometrical relation as decreed by the following fundamental equation of the equilateral or rectangular hyperbola. See pages 69-77

$$\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1$$

If " a " and " b " are assumed to be unity, then $X^2 - Y^2 = 1 = \cosh^2 \theta - \sinh^2 \theta$.

Again sinh and cosh cannot be and are not intrinsically defined by an expression which only conveniently provides their precise numerical evaluation.

A GRAPHICAL ILLUSTRATION

An equation, in general, is a "shorthand" statement of a simple and basic truth. An equation makes explicit in symbolic terms a fact which otherwise might need many words for satisfactory expression. If actual numerical values for the quantities embodied in the equation are invoked to illustrate the meaning of the equation, then it is likely that the equation takes on a bit more of meaning and reality.

Inspection of the accompanying table of numerical values is therefore in order. The values of cosh and sinh were taken from one table and the values of e^x and e^{-x} were taken from another table. With a bit of arithmetic, one sees readily that the equations involved are numerically correct.

TABLE

X	Cosh X	Sinh X	Cosh $X + \sinh X$ $= e^x$	Cosh $X - \sinh X$ $= e^{-x}$
.00	1.000	.000	1.000	1.000
.25	1.031	.2526	1.284	.779
.50	1.128	.5211	1.649	.606
.75	1.295	.8223	2.117	.472
1.00	1.543	1.175	2.718	.368
1.25	1.888	1.602	3.490	.286
1.50	2.352	2.129	4.481	.223
1.75	2.964	2.790	5.755	.174
2.00	3.762	3.627	7.389	.135
2.25	4.797	4.691	9.488	.105
2.50	6.131	6.050	12.182	.082
2.75	7.853	7.789	15.643	.064
3.00	10.07	10.02	20.09	.05
3.50	16.57	16.54	33.11	.03
4.00	27.31	27.29	54.60	.02

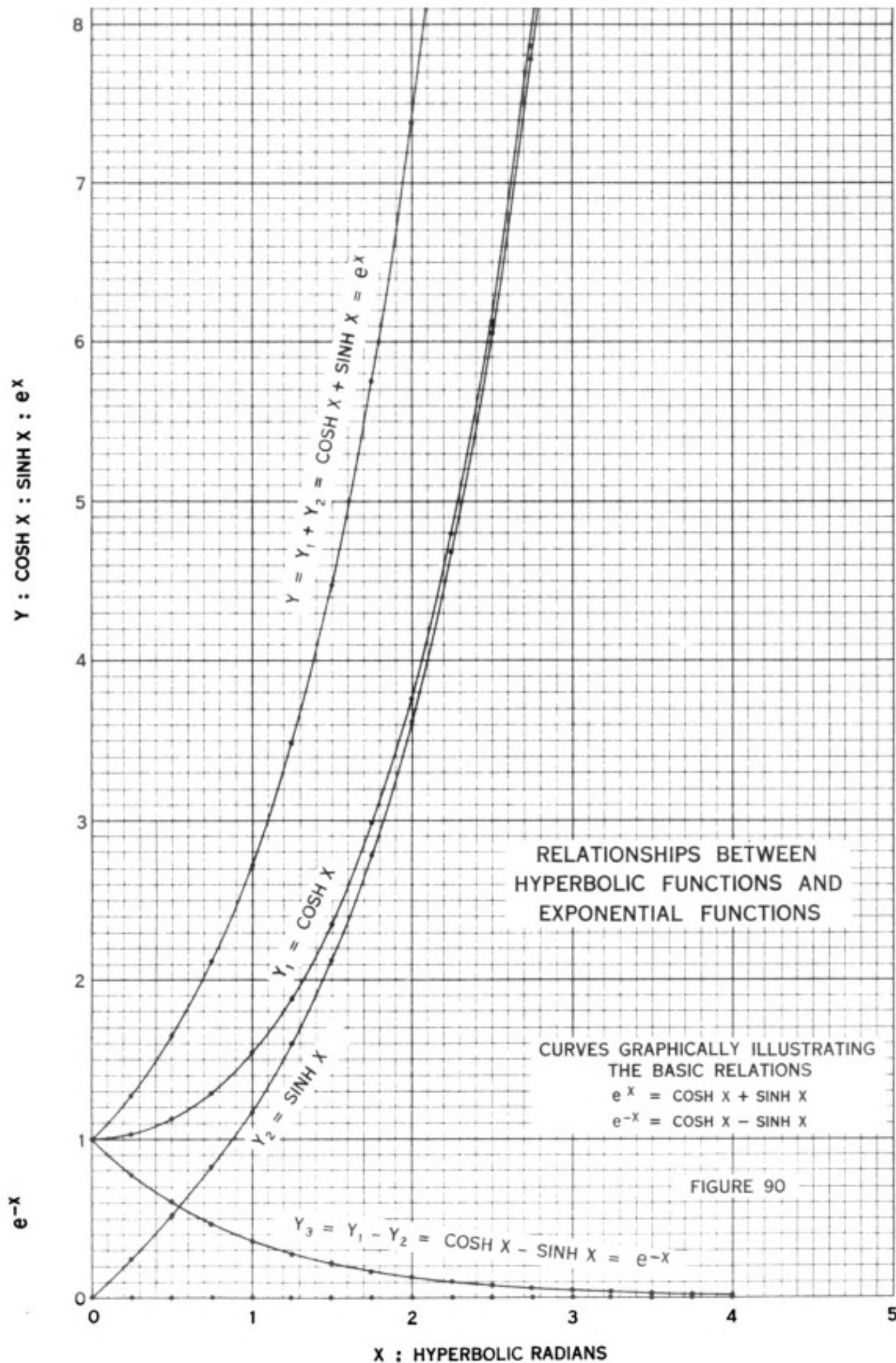
If the numerical values of an equation are plotted and curves drawn, one gets a further realistic appreciation of the equation. Graphical illustrations are always helpful and impressive. Cosh X , sinh X , e^x and e^{-x} can all be plotted to provide four curves. Figure 90 The numerical values used are those of the accompanying table. It is to be seen by inspection that the value of every Y on the e^x curve for values of X is the sum of the corresponding Y values of the cosh and sinh curves. Likewise the values for every Y on the e^{-x} curve is the difference of the corresponding Y values of the cosh and sinh curves.

In discussing and portraying, by curves, the decreasing amplitude of a sine curve on page 115, Figs 46, 47, it was brought out that the product of the corresponding values of two curves provided the values of a third curve. The $Y = 20 \sin \theta$ values were multiplied by values in the curve $Y_1 = 20 e^{-.20\theta}$ to produce, as it were, the equation of the decaying oscillatory curve. In like manner, the curve of a speech or musical sound can be determined by adding the Y values of a number of sine waves of different frequencies. This doctrine is the Fourier series doctrine. In the present case, the values of two curves are added and subtracted to provide a third and fourth curve.

The use of a bit of arithmetic applied to the table shows readily that cosh X equals $\frac{e^x + e^{-x}}{2}$ and that sinh $X = \frac{e^x - e^{-x}}{2}$.

$$\sinh .5 = \frac{1.649 - .606}{2} = .5211$$

$$\cosh .5 = \frac{1.649 + .606}{2} = 1.128$$



This relation cannot, however, be noted as readily by inspection of the curves as can $e^x = \cosh X + \sinh X$ and $e^{-x} = \cosh X - \sinh X$.

It is important to point out that the numerical values of $\cosh$ and $\sinh$ come out of the intrinsic nature and geometry of an equilateral hyperbola. In its own self, the hyperbola has no direct connection with exponentials and geometric progression and retrogression. However, it is impressive to realize that the sum and difference of $\cosh$ and $\sinh$ actually provide two exponential equations which pertain to geometric progression and retrogression. It is a striking coincidence.

It is interesting to observe that the e^x curve in the accompanying chart reaches its highest value of 9.98 at an x value of 2.3. At this value of x , $\cosh$ and $\sinh$ have a value of 5.04 and 4.94 respectively. At a value of 3 for x , the $\cosh$ is 10.07 and $\sinh$ is 10.01. The sum of these (20.08) would be a point on the e^x curve high above the top of the chart.

Geometric growth and decay is a basic fact in many natural phenomena. To have the $\cosh$ and $\sinh$ interject themselves into the scene is surprising and impressive.

EXPONENTIALS APPLIED TO CIRCULAR SINES AND COSINE

In some text books, the statement is made that "circular trigonometric functions do not need to be given a geometrical interpretation". The writer does not agree with this. Sines and cosines are inextricably associated with a circle and its equation

$$\begin{aligned} X^2 + Y^2 &= R^2 \\ \left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 &= 1 \\ \cos^2 \theta + \sin^2 \theta &= 1 \end{aligned}$$

One cannot in the beginning of exposition invoke any doctrine other than simple algebra and simple basic geometry which are readily understood and appreciated by the young student. $\sin X$ and $\cos S$ are not "just functions" as some specialists say. They are functions which exist because of the actual geometry of a circle with readily interpretable portrayals. $\sinh$ and $\cosh$ are not "just functions". They come out of simple premises based on the simple formula of a simple curve, the equilateral hyperbola. They are rationalized by considering the geometry of this equilateral hyperbola (Page 77).

Following the above quoted statement, a second statement oftentimes follows "Circular trigonometric functions can be defined by an algebraic exponential form". Again the writer disagrees with this interpretation of the word "define". What algebraic exponential forms provide is not a definition but a mathematical "tool" to provide vectorial portrayal and interpretation.

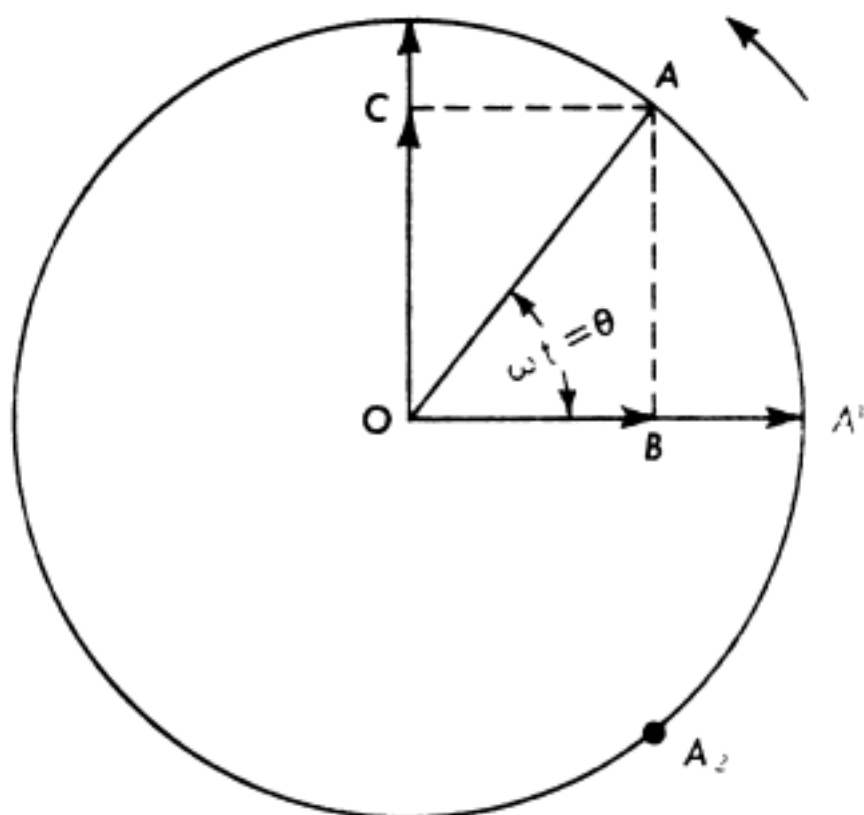
The geometric and graphical approach combined with the basically simple circular and hyperbola trigonometric relations are most valuable in the learning process. They contribute to understanding. The vector, the vector operator, exponential doctrines and techniques are excellent but one must know and understand something about the simple and basic physical and geometric relations before he can use the special tools of mathematics.

The exponential expressions for sine and cosine are vectorial expressions; vectorial additions and subtractions. They, unlike their series expressions, do not provide an expression which per se permits of actual numerical calculation. Of course they can be used to obtain numerical values by knowing numerical values of sine and cosine and then using the Pythagorean theorem.

The objective at this point in the discussion is to analyze, understand, and get a "feel" for the expressions

$$\begin{aligned} \cos x &= \frac{e^{jx} + e^{-jx}}{2} \\ \sin x &= \frac{e^{jx} - e^{-jx}}{2j} = -j \left(\frac{e^{jx} - e^{-jx}}{2} \right) \end{aligned}$$

as contrasted to the same sort of expressions for $\sinh x$ and $\cosh x$ previously given on page 153.



GRAPHICAL REPRESENTATION OF
COUNTER-CLOCKWISE ROTATING VECTOR
FIGURE 91

In these expressions for $\sin X$ and $\cos X$, j is the often used vector "operator" indicating a counterclockwise rotation of $+90^\circ$ or $+\frac{\pi}{2}$ radians. Sometimes i is used instead of j . For a clockwise rotation of $-\frac{\pi}{2}$, $-j$ or $-i$ is used.

For a rotation of $+180^\circ$ or $+\pi$ radians, the vector operator is -1 . The j or i is equivalent to $\sqrt{-1}$. Two $+\frac{\pi}{2}$ counterclockwise rotations in succession are indicated by i^2 or j^2 or -1 . Minus one (-1) has long been used in arithmetic and algebra to indicate the opposite direction which is equivalent to 180° or π rotation.

SOME SIMPLE AND BASIC VECTOR ANALYSIS

It is again to be stated and emphasized that circular trigonometric functions are recurrent, cyclic and periodic functions. They enter the picture because certain quantities like voltage and current are produced by an armature winding rotating in a circle. So let's go back to simple fundamentals.

OA in the accompanying Figure 91. is the graphical representation of some physical or electrical quantity or factor with a name. *e. g.* voltage. It is called a vector and the vector is assumed to rotate counterclockwise. The factor represented by OA may be a voltage E which is supplied by an electric generator and is varying sinusoidally in the output supply of the generator.

Assume the factor E or OA , equal to the radius of the circle, rotates counterclockwise through an angle θ . The angular velocity of OA or E is ω (omega) in radians per second. Therefore θ (the angle) $= \omega t$ (radians). Let's assume that $OA = E = \text{unity (1)} = OA$.

As per fundamentals

$$\text{circular } \sin \theta = \frac{AB}{AO}$$

$$\text{circular } \cos \theta = \frac{OB}{AO}$$

Since OA was assumed to be unity (1)

$$\sin \theta = AB \text{ (pictorially, graphically)}$$

Likewise

$$\cos \theta = OB \text{ (pictorially, graphically)}$$

Now for sake of argument and arrival at the immediate objective we set for ourselves, the simple elements of vector analysis, assume that I myself wish to go physically and personally from point O to point A . I could go right out along a radius itself and I would arrive at point A . The distance or numerical value of my journey would be the distance OA , namely the radius.

However, to get from point O to point A , I could go along the X coordinate axis to point B , then make a left turn of 90° or $\frac{\pi}{2}$ radians and go straight up to point A . I have taken a longer path by taking two separate routes, but I have arrived at the same destination.

Now from the standpoint of vector analysis or just vectors, one says $OA = OB + BA$.

$OB + BA$ does not represent the actual numerical distance OA , but nevertheless one can say vectorially $OA = OB + BA$. This is a vectorial addition.

Now OB is $\cos \theta$ and BA is $\sin \theta$.

So $OA = \cos \theta + \sin \theta$ (vectorially).

$OA^2 = \cos^2 \theta + \sin^2 \theta$ (numerically).

It is to be recalled that I made a left hand turn, a counterclockwise turn of 90° or $\frac{\pi}{2}$ at point B .

Therefore I can write symbolically to emphasize the fact that I did make the $\frac{\pi}{2}$ turn at point B .

$$OA = OA_1 (\cos \theta + j \sin \theta).$$

For any and every point or position on the circle as point A started along its circular path and finally up to point F , we could write as a vectorial expression.

$$OA \text{ (wherever point } A \text{ is)} = OA_1 (\cos \theta + j \sin \theta)$$

A different angle (θ) of rotation of the radius vector OA_1 would be involved of course for the different places that point A might be.

Even if OA_1 the original rotating vector was not unity (1) but was E , a voltage then

$$\begin{aligned} OA &= (\cos \theta + j \sin \theta) E \\ &= (\cos \omega t + j \sin \omega t) E = E \cos \omega t + j E \sin \omega t \end{aligned}$$

The simple and straightforward truth in the above expression or formula is oftentimes called DeMoivre's theorem. Like all great truths, it is a simple truth expressed in the symbolism of mathematics.

TWO VECTORS

The rotating vector E can therefore at any instant of time be readily thought of as composed of two vectors of the vector sum (not numerical) of two vectors. One of these two vectors is $E \cos \omega t$ which varies (as the original vector OA rotates) along the X coordinate axis. The second of the two vectors is $E \sin \omega t$ which varies (as the original vector OA_1 rotates) along the Y coordinate axis.

Since one must make the left hand turn, symbolized by the vector operator j , one says the second varying vector is $j E \sin \omega t$.

In these vectorial and mathematical endeavors, the X coordinate axis varying vector ($E \cos \omega t$) is said to vary along the "real" axis. The Y coordinate axis varying vector ($j E \sin \omega t$) is said to vary along an "imaginary axis". The Y axis is not really "imaginary" in the ordinary sense of the word. The j or the $\sqrt{-1}$ is a piece of "vector machinery" to indicate a 90° left hand turn and the $\sqrt{-1}$ is sort of an "imaginary quantity" from the standpoint of ordinary algebra and squares and square roots. Hence the term "imaginary" may be justified. In other words, the $\sqrt{-1}$ has no reality in terms of ordinary mathematical doctrine and is an "imaginary" or hypothetical term; a useful and interpretive tool however. The $j E \sin \omega t$ is just as actually real as $E \cos \omega t$ but some word had to be used to remind one of the left hand turn in vector notation so (unfortunately it may seem) the Y axis is called the "imaginary" axis. It might have been better, when this kind of analysis was first developed, that the simple terms, horizontal axis and vertical axis would have been adopted.

It is again to be pointed out and emphasized that the actual numerical value of OA is not ($E \cos \omega t + j E \sin \omega t$) but is a vector sum expressed in vector language. Of course, the vector expression can be used to obtain numerical values by knowing actual values of $\sin \omega t$ and $\cos \omega t$ and then using the Pythagorean theorem.

As stated above, the rotating vector E can be thought of as the net result of two vectors of equal amplitude; one vector ($E \cos \omega t$) varies along the so called real X coordinate axis and the other vector ($j E \sin \omega t$) varies along the so called imaginary Y axis.

The X axis vector ($E \cos \omega t$) is a maximum (E) when vector OA (E) starts out in position OA_1 since the angle θ or ωt at that instant is zero (0). The cosine of zero (0) angle is unity (1). Therefore the vector $E \cos \omega t$ leads the $j \sin \omega t$ vector by 90° or $\frac{\pi}{2}$ since the latter is zero when the former is a maximum at t (time) = 0. At angle θ = zero ($t = 0$), the $j E \sin \omega t$ is zero since sine of zero (0) angle is zero (0).

CLOCKWISE ROTATION

If vector OA_1 had started off rotating from its original position along X axis in a clockwise fashion rather than counterclockwise then obviously in a vector language or notation.

$$OA = E \cos \omega t - j \sin \omega t$$

In the vector notation, $+j$ denotes or prescribes a "left hand turn" of $\frac{\pi}{2}$ radians (90°) and $-j$ prescribes a right hand turn of $\frac{\pi}{2}$ radians.

POLAR COORDINATES

In approaching and analyzing the straightforward expression (DeMoivres' theorem) developed out of analysis of accompanying figure ($OA = E \cos \omega t + j E \sin \omega t$), we said one could go from 0, the origin, to point A by two routes or two methods.

The first method or scheme was to go directly out on radius vector, which had been rotated an angle θ or ωt , to the destination, point A . The second method was to go via OB , make a left turn (do a j as it were) and go up to point A .

Symbolically we could appropriately say that the first method could be indicated

$$OA = OA_1 | \theta | = E | \theta | = E | \omega t |$$

The two vertical lines about θ and ωt are a symbol of rotation of OA to offset the possible meaning of sheer multiplication. Sometimes the symbolism is a bit different, namely OA/θ or E/θ . Both symbolisms indicate the same procedure; they are a sort of code and "short hand" language.

These latter symbolic expressions could be called polar coordinate notation. They say; we have a radius vector E and it has rotated counterclockwise θ or ωt . If we do what the polar expression suggests, point A_1 has arrived at position A and then one is to go from 0 to A , by walking straight out on the rotated radius vector to the end of line, namely position A .

This doctrine is a polar coordinate doctrine. The $E (\cos \omega t + j \sin \omega t)$ is another doctrine, namely a vectorial doctrine. Both doctrines mean the same thing; they achieve the same result.

VECTOR CONSIDERATIONS CONTINUED

In the case of hyperbolic functions, the series expansion doctrine provided a series expression for $\sinh X$ and $\cosh X$. From straightforward considerations coming out sheer arithmetical addition and subtraction of the three series expressions, it was readily resolved that

$$e^x = \cosh X + \sinh X$$

$$e^{-x} = \cosh X - \sinh X$$

$$\text{Also } \sinh X = \frac{e^x - e^{-x}}{2}$$

$$\cosh X = \frac{e^x + e^{-x}}{2}$$

All of these above expressions provide actual numerical values. They are not vectorial additions and subtractions. The vector operator j or $-j$ is not involved. It was emphasized that there are no complete 2π cyclic rotations of a vector counterclockwise in the case of hyperbolic path as these rotations do exist for circular paths to provide circular sines and cosines.

But as a matter of further analysis of the circular rotation venture, let's explore the matter a bit further. Let's explore and maneuver a bit in the realm of the expansion series for expansion series for e^x , e^{-x} , $\sin X$ $\cos X$, and the vector operator j .

It would be entirely valid and tenable to have the X in the e^x be an angle (θ) since angle is a pure numeric as pointed out earlier. Furthermore, $\cos \theta$ and $\sin \theta$ or $\cos \omega t$ and $\sin \omega t$ are sheer numerics.

In De Moivre's vector theorem as previously given

$$\begin{aligned} OA &= E (\cos \omega t + j \sin \omega t) \\ &= E (\cos \theta + j \sin \theta) \end{aligned}$$

The OA_1 is vector and a voltage but the $\cos \theta$ and $\sin \theta$ are pure numbers involving another pure number, namely the angle θ . Voltage is not a vector strictly speaking but it can be portrayed by a straight line as if it were a vector.

A BIT OF SUMMARY

As per polar coordinate doctrine and notation, we put down E/θ or $E/\omega t$ to symbolize that a vector E was to be rotated counterclockwise by an angle, θ or ωt . The numerical value of the original vector E is not changed thereby and is not actually multiplied by anything. The notation $/\theta$ is merely a symbol or a code to indicate a rotation. The $/\theta$ is an "operator" just as $\sqrt{-1}$ is.

Another "operator" to indicate rotation of vector E through an angle θ is obtained from De Moivre's theorem, namely $(\cos \theta + j \sin \theta)$.

So far in this discussion, we have three symbolic notations to indicate rotations. The $\sqrt{-1}$ or j is an operator which indicates a rotation of 90° in a counterclockwise fashion. The j has no numerical value in the ordinary sense of the word. When one writes $j \sin \theta$, one does not mean that $\sin \theta$ has been numerically multiplied by something.

The second operator to indicate rotation is the simple symbol $/\theta$.

The third operator is the $(\cos \theta + j \sin \theta)$ and it indicates exactly the same procedure as $/\theta$. All three operators mean that the original vector E has been rotated counterclockwise by an angle θ ; they do not mean that the original vector E has been multiplied by anything to change its numerical value.

THE EXPONENTIAL "OPERATOR"

Now for purposes of mathematical exploration and a hoped for valuable and important tool, we can search for still another "operator." One says, let us take advantage of the exponential technique and doctrine.

As a perfectly valid and tenable procedure we can write $/\theta$ as $e^{j\theta}$ for a counterclockwise rotation and $e^{-j\theta}$ for a clockwise rotation to operate on the vector E . Again it is to be pointed out that in the expression $e^{j\theta}$, the ordinary process of multiplication of E by $e^{j\theta}$ is not implied but merely that the vector E is to be subjected to a rotation. The new code is $e^{j\theta}$; $e^{j\theta}$ is another type of "vector operator."

It is to be particularly noted that the amount of rotation is only θ or ωt even though there is a "j" involved which "j" usually of itself indicates a 90° or $\frac{\pi}{2}$ rotation. In this exponential operator doctrine, a rotation of 90° or $\frac{\pi}{2}$ is indicated by $e^{j\frac{\pi}{2}}$.

The "j" of the $e^{j\theta}$ operator merely identifies it as a rotation indicator and to distinguish it from e^θ as such where θ is an exponent of "e". The $j\theta$ of $e^{j\theta}$ is not to be regarded as an exponent of "e" in the conventional meaning and sense of the word. The $j\theta$ of $e^{j\theta}$ is sometimes called an imaginary exponent.

The "e" is the same "e" encountered and mastered before; it is the Napierian base; oftentimes written as ϵ , epsilon.

To summarize we can say; previously the radius vector was OA ; but now, for sake of exploration, OA (E) is expressed as $E e^{j\omega t}$ and $E e^{-j\omega t}$ to symbolize the rotation of the radius vector E through an angle ωt in a counterclockwise fashion and in a clockwise fashion respectively. Here again, the process of multiplication in the usual sense is not implied.

So let's write the expansion series for e^{jx} or $e^{j\theta}$ which may provide some useful information. We will use the expansion series for e^x (page 227) as given on previous page, and substitute jx for x and proceed in the conventional manner, just as if jx were a genuine exponent.

$$e^{jx} = 1 + jx + \frac{j^2 x^2}{2} + \frac{j^3 x^3}{3} + \frac{j^4 x^4}{4} + \frac{j^5 x^5}{5} + \frac{j^6 x^6}{6} + \dots$$

Now since the vector operator j is $\sqrt{-1}$, j^2 is -1 , j^3 is $-j$, j^4 is 1 , j^5 is j and j^6 is -1 , the e^{jx} becomes

$$e^{jx} = 1 + jx - \frac{x^2}{2} - \frac{jx^3}{3} + \frac{x^4}{4} + \frac{jx^5}{5} - \frac{x^6}{6}$$

It now is seen that e^{jx} has some terms in its series expansion which have a "j" and some which do not have a "j".

Separating the "j" terms from the non "j" terms gives the following

$$e^{jx} = 1 - \left(\frac{x^2}{2} + \frac{x^4}{4} - \frac{x^6}{6} \right) + j \left(x - \frac{x^3}{3} + \frac{x^5}{5} \right)$$

Now it is striking to observe that the first parentheses is actually the series expansion for the circular cosine and the second parenthesis is actually the series expansion for the circular sine.

$$\text{Therefore } e^{jx} = \cos X + j \sin X = e^{j\theta} = \cos \theta + j \sin \theta.$$

We can now express the De Moivre vector theorem in an exponential vector form;

$$OA = E e^{j\theta} = E e^{j\omega t} = E (\cos \omega t + j \sin \omega t)$$

Before by vector addition, it was simply established that OA was equal to $\cos x + j \sin x$. Now we have reestablished as it were, that it was valid and legitimate to assume that OA_1 could be replaced by $E e^{jx}$.

But we substituted jx for x in e^x as a matter of valid and tenable exploration and so we have

$$\begin{aligned} |E| \angle \theta &= E (\cos \theta + j \sin \theta) \\ &= E \cos \theta + j E \sin \theta \end{aligned}$$

The vertical lines alongside E means that it is the numerical value of E that is involved and the $\angle \theta$ is to indicate the angle through which it was rotated. There is no multiplication implied.

For clockwise rotation $E e^{-j\theta} = E (\cos \theta - j \sin \theta)$

TWO VECTORS

So we have two vectors $E e^{j\theta}$ and $E e^{-j\theta}$ or $E e^{j\omega t}$ and $E e^{-j\omega t}$. The vector E is operated on by $e^{j\omega t}$ to rotate counterclockwise and is operated on by $e^{-j\omega t}$ to rotate clockwise. The outstanding result is that now we have expressed the "operator" $(\cos \omega t + j \sin \omega t)$ in terms of an exponential "operator," $e^{j\omega t}$. The $e^{-j\omega t}$ is equivalent to $(\cos \omega t - j \sin \omega t)$

It will be helpful and significant to recognize that if rotation is 90° or $\frac{\pi}{2}$ then the exponential operator becomes

$$e^{j\frac{\pi}{2}} = \cos \frac{\pi}{2} + j \sin \frac{\pi}{2} = 0 + j = j$$

This checks as it were; our exponential approach was valid.

$$\text{Also } e^{-j\frac{\pi}{2}} = \cos \frac{\pi}{2} - j \sin \frac{\pi}{2} = 0 - j = -j$$

Also, it will be worth while to consider a rotation greater than $\frac{\pi}{2}$.

$$\begin{aligned} e^{j(\theta + \frac{\pi}{2})} &= e^{j\theta} e^{j\frac{\pi}{2}} = j e^{j\theta} \\ e^{j(\theta - \frac{\pi}{2})} &= e^{j\theta} e^{-j\frac{\pi}{2}} = -j e^{j\theta} \end{aligned}$$

FOUR "OPERATORS"

Finally then we have four "operators" which are equivalent namely $\angle \theta$, j , $(\cos \theta + j \sin \theta)$ and $e^{j\theta}$. They are four different codes which mean the same. They all indicate and symbolize a counterclockwise rotation of the vector. It is to be particularly pointed out and emphasized that

$$\begin{aligned} e^{j\theta} &= \cos \theta + j \sin \theta \\ e^{-j\theta} &= \cos \theta - j \sin \theta \end{aligned}$$

are exponential and vectorial expressions for addition and subtraction and are not arithmetical expressions for addition and subtraction.

To be sure, these vectorial expressions can be used to get numerical values by finding numerical values of $\cos \theta$ and $\sin \theta$ and then invoking the Pythagorean theorem.

The above vector expressions make plain that

$$e^x \text{ does not equal } \cos x + j \sin x \text{ numerically}$$

$$\text{and } e^{-x} \text{ does not equal } \cos x - j \sin x \text{ numerically.}$$

In the case of hyperbolic functions, the equations are not vectorial since no "operators" are involved.

$$e^x \text{ does actually equal } \cosh x + \sinh x, \text{ numerically}$$

$$e^{-x} \text{ does actually equal } \cosh x - \sinh x, \text{ numerically}$$

See Figure 90, page 229.

In general among several things we have achieved so far, we have an exponential expression for hyperbolic cosine and hyperbolic sine and are ready to work on the idea for circular cosine and sine.

CONTINUED EXPLORATION ABOUT SIN X AND COSINE X

Since it has been logically and simply developed that as a matter of vector and exponential language that

$$e^{j\theta} = \cos \theta + j \sin \theta$$

$$e^{-j\theta} = \cos \theta - j \sin \theta$$

we can, as a matter of further exploration and maneuvering, add $e^{j\theta}$ and $e^{-j\theta}$ in a manner similar to that on page 350 for sinh and cosh.

If we do an addition, we readily get

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

If we subtract $e^{-j\theta}$ from $e^{j\theta}$, we readily get

$$\sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j} = -j \left(\frac{e^{j\theta} - e^{-j\theta}}{2} \right)$$

It again is to be pointed out and emphasized that these are vector equations, since they involve j .

In passing, it might be well to stop a while and see why the j in denominator becomes a $(-j)$ in the numerator of above two expression for $\sin \theta$.

The operator j is conventionally prescribed as $\sqrt{-1}$ as pointed out before.

Then $j^2 = -1$ and $-j^2 = -(-1) = 1$.

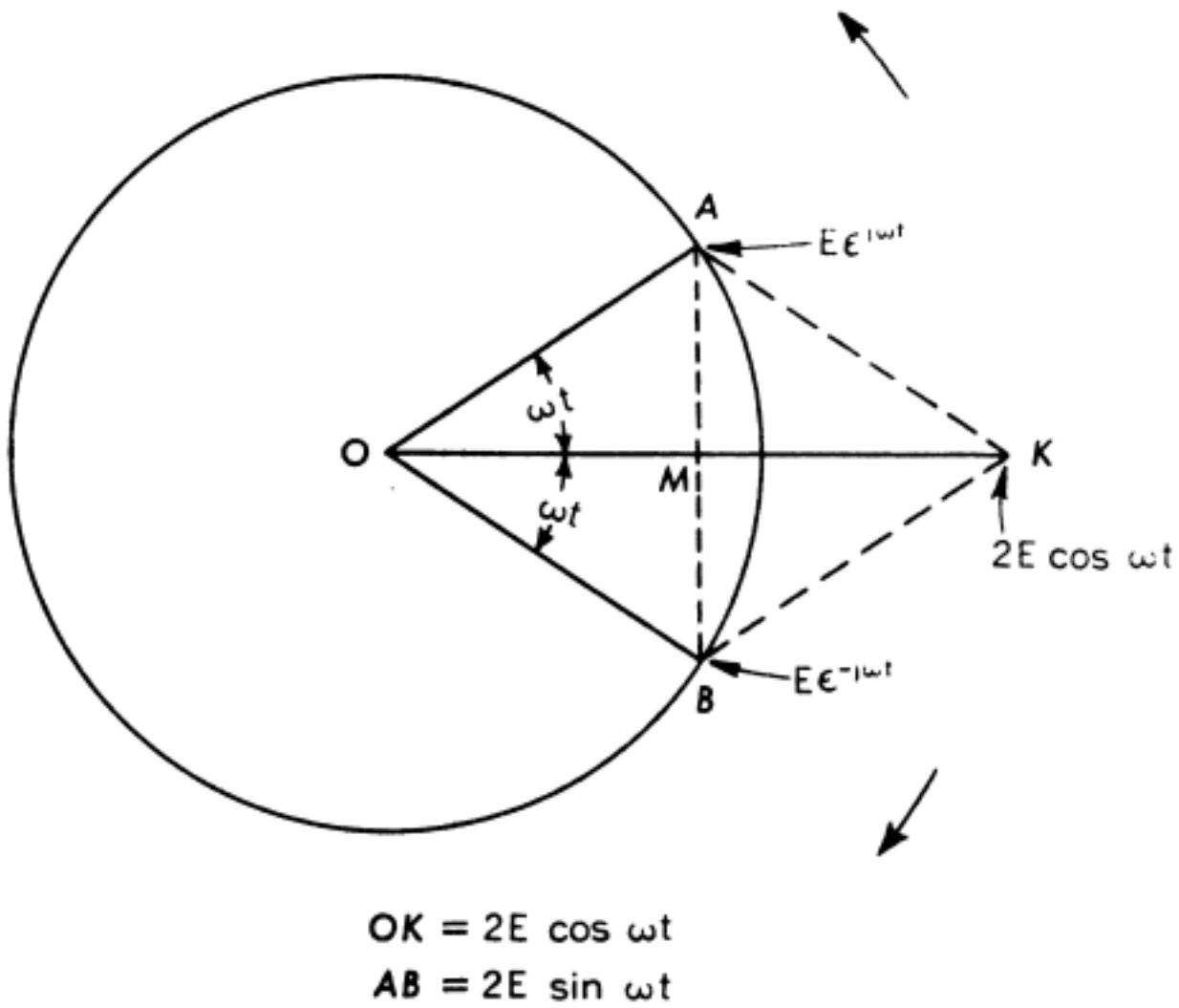
Now $\frac{e^{j\theta} - e^{-j\theta}}{2j}$ can be thought of as having a one (1) in front of it.

$$1 \left(\frac{e^{j\theta} - e^{-j\theta}}{2j} \right) = -j^2 \left(\frac{e^{j\theta} - e^{-j\theta}}{2j} \right) = -j \left(\frac{e^{j\theta} - e^{-j\theta}}{2} \right)$$

In general again, we have achieved an exponential and vectorial expressions for the circular functions, sine and cosine. Again it is emphasized that they are not arithmetical expressions.

GRAPHICAL REPRESENTATION AND INTERPRETATION OF TWO ROTATING VECTORS

In accompanying Figure 92, the radius vector $E e^{j\omega t}$ is shown as having been rotated counterclockwise through an angle θ , or ωt . During the same time, it is assumed that the radius vector has been rotated clockwise. It is shown as $E e^{-j\omega t}$. In both cases the ω (omega) is the angular velocity in radians per second; as usual $\theta = \omega t$.



GRAPHICAL REPRESENTATION OF
TWO ROTATING VECTORS

FIGURE 92

By the simple and ordinary process of adding vectors as was done with concurrent forces in elementary physics, the vector sum of the two equal radii vectors at a particular instant of time (t) is the line OK on figure.

As per figure, point M is the point of intersection of the two diagonals of the parallelogram $OAKB$.

As per original simple and basic ideas about sines and cosines in figure, OM , is $E \cos \omega t$ and OK is $2E \cos \omega t$. Likewise AM is $E \sin \omega t$ and AB is $2E \sin \omega t$. A $(j$ and a $(-j)$ might accompany the two halves of AB . $\cos \omega t = \frac{1}{2} OK$

But OK is the vector sum at a particular instant of time of the two rotating vectors $Ee^{j\omega t}$ and $Ee^{-j\omega t}$.

Finally therefore it is graphically seen that

$$OM = \frac{OK}{2} = E \cos \omega t$$

Likewise

$$= \frac{Ee^{j\omega t} + Ee^{-j\omega t}}{2}$$

$2 E \sin \omega t$ is graphically seen to be equal to AB .

AM is, as per original doctrine, equal to $\sin \theta$ and of course BM is equal to $-jE \sin \theta$ where $j^{-\theta}$ ($j^{-\omega t}$) denotes a clockwise rotation. In other words $AM = -j BM$.

So it is easy to see

$$\begin{aligned} AM &= -j BM = \frac{AB}{2} \\ &= -jE \sin \omega t = -j \left(\frac{Ee^{j\omega t} - Ee^{-j\omega t}}{2} \right) \end{aligned}$$

Graphically the $-j E \sin \omega t$ is represented by BM in the drawing.

VECTOR REPRESENTATION OF A DAMPED OR ATTENUATING SINUSOIDAL WAVE

In the pages 115, 118 there was presented a discussion, with curves, of decreasing amplitude of tuning forks and alternating current or voltage.

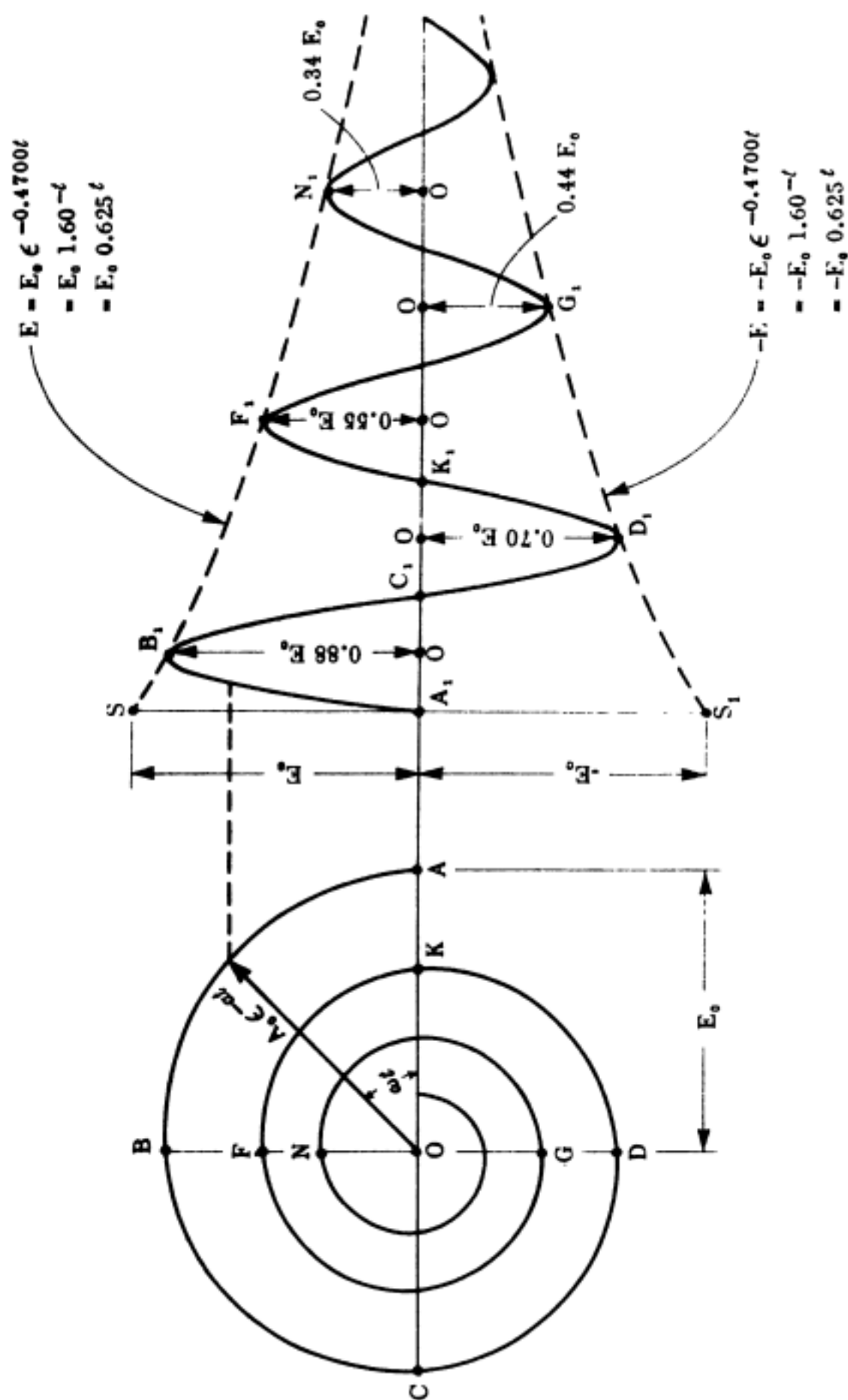
It is in order to now approach this problem from a different point of view, namely the vector point of view. The equation $E = E_0 e^{j\omega t}$ represents a counterclockwise rotating vector. If one would plot a curve with E (voltage) on the Y axis and time (t) on the X axis, one would get a sine curve as per page 27, Figure 16 and the damped curves of pages 115, 118, Figs 46, 47. In one case the maximum amplitudes, E , (or I) remain constant in magnitude. In the case of attenuation, the E (or I) decreases as illustrated in the two equations of the latter named pages.

$$I_1 = 20e^{-.02\theta} \text{ and } I_2 = 20e^{-.2\theta}. \text{ The } \theta \text{ is } \omega t.$$

One can now write the equation of the damped sine wave as follows.

$$\begin{aligned} E &= E_0 e^{-\alpha t} e^{j\omega t} \\ &= E_0 e^{-\alpha t} (\cos \omega t + j \sin \omega t) \\ &= E_0 e^{-\alpha t} \cos \omega t + j E_0 e^{-\alpha t} \sin \omega t \end{aligned}$$

The changing and decreasing vector E (maximum value E_0) is at all times the vector sum of two vectors; one along the axis of reals (X axis) and the other along the axis of imaginaries (Y axis). The magnitudes of these two components vary relatively with elapsed time as the sine and cosine always do. Furthermore the magnitudes of the two components are in addition controlled by the shrinking factor $e^{-\alpha t}$ in their individual expressions. Their numerical values at every instant of time (except at $t = 0$) are less than they would be with no damping.



VECTOR REPRESENTATION OF A DAMPED SINUSOID

FIGURE 93

The situation is represented in the accompanying drawing, Figure 93. When $t = 0$, the rotating vector E is equal to $E_0 = OA$ in the spiral curve, since the $e^{-\alpha t} \cos \omega t$ of its X axis component is 1 because $e^0 = 1$ and $\cos 0$ is 1. The initial terminus of the rotating vector E_0 at $t = 0$ is at point A in drawing. On the other hand, the Y axis component $jE_0 e^{-\alpha t} \sin \omega t$ is zero at $t = 0$ because $e^{-\alpha t} = e^0 = 1$ and $\sin 0 = 0$. As per drawing, the initial point on the damped curve is at point A_1 .

As time continues, the rotating vector E shrinks in magnitude as per the spiral path of its terminus since as said before both components change relatively. At $t = T/4$ or at an ωt of $\frac{\pi}{2}$ (90°), E has a magnitude of OB . At $t = T/2$ and an ωt of (180°), $E = 0$ because terminus of vector is at point C .

At $t = \frac{3T}{4}$ and an ωt of $\frac{3\pi}{2}$ (270°), $E = OD$ since terminus of rotating vector has moved counterclockwise around the spiral to point D . At $t = T$ and an ωt of 2π , the terminus of the vector is at K .

It is to be observed that the ratio of the successive maximum values of the amplitude of the curve is a constant. Each succeeding amplitude is less than the previous amplitude by the same amount which is numerically less than unity. The principle of geometric retrogression prevails; the natural law of decay is "on the job."

In making the accompanying drawing (Fig. 93) the logarithmic decrement constant was, for illustrative purposes, arbitrarily selected as .47. This means that $e^{-.47} = 1.60 = r$ and that $e^{-.47} = .625 = r^{-1}$. As per drawing $E_0 = OA$ of the spiral and vertical line SA_1 of the curves. The time (t) is 0.

At $t = T/4 = 1/4$, the decay expression is $e^{-.47}$ or $e^{-.118}$ with the result that $OB_1 = .88E_0$. At $t = 3/4$, $OD_1 = e^{-3/4 \cdot .47} E_0 = e^{-.353} E_0 = .70E_0$. At $t = 5/4$, $OF = OF_1$ becomes $e^{-.588} E_0 = .55E_0$. At $t = 7/4$, OG_1 becomes $e^{-.725} E_0 = .44E_0$.

The above calculations provide the ratio of amplitudes:

$$\begin{aligned} \frac{OB_1}{OF_1} &= \frac{.88E_0}{.55E_0} = 1.60 = r & \frac{OF_1}{OB_1} &= .625 = r^{-1} \\ \frac{OD_1}{OG_1} &= \frac{.70E_0}{.44E_0} = 1.60 = r & \frac{OG_1}{OD_1} &= .625 = r^{-1} \end{aligned}$$

If one desired to consider only the ratio of successive positive and negative amplitudes as well as the ratio of successive positive and successive negative amplitudes rather than intermediate values of E involving e^α , it can be done readily in terms of r^2 , r , $\sqrt[3]{r}$ and $\sqrt[4]{r}$.

$$\begin{aligned} OB_1 &= rOF_1 = 1.60OF_1 = \sqrt{1.60} OD_1 = 1.26OD_1 \\ &= r^2 ON_1 = 2.56 ON_1 \end{aligned}$$

$$OD_1 = 1.60OG_1 = \sqrt{1.60} OF_1 = 1.26OF_1$$

$$OF_1 = \sqrt{1.60} OG_1$$

$$E_0 = AO = A_1S = \sqrt{1.60} OB_1 = 1.12OB_1$$

Again it is to be stated that undamped sinusoidal functions can be expressed as the vector sum of two vectors. Damped sinusoidal functions can be expressed also as the vector sum of two vectors with an added negative exponential term indicating the damping.

A BIT OF SUMMARY

The validity, simplicity and interpretation of the counterclockwise rotation vectorial "operator" ($\cos \theta + j \sin \theta$) is readily grasped by a graphical analysis and portrayal. For a clockwise rotation, the "operator" becomes ($\cos \theta - j \sin \theta$).

It has been developed by understandable and logical step by step arguments that ($\cos \theta + j \sin \theta$) is expressible by another "operator"; namely the exponential operator $e^{+j\theta}$. The counterparts are ($\cos \theta - j \sin \theta$) and $e^{-j\theta}$.

By simple additions and subtractions of the understandable equations expressed in expansion series form, the exponential and vectorial expressions for $\sin \theta$ and $\cos \theta$ are a natural and logical result; namely

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

$$\sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j} = -j \frac{e^{j\theta} - e^{-j\theta}}{2}$$

One final statement is to be made in conclusion which is a repetition of a statement previously made.

It is sometimes said that circular sines and circular cosines do not need to be given a geometrical interpretation and that they can be defined by algebraic exponential forms. The above expressions do not constitute a very lucid definition. The above use of the word "define" is not a valid use. Basically sines and cosines in all respects come naturally and simply from the equation and the geometry of a circle

$$\left(\frac{X}{R}\right)^2 + \left(\frac{Y}{R}\right)^2 = 1 = \cos^2 \theta + \sin^2 \theta$$

All of the mathematical exploration that has been discussed was predicated on the above basic equation for a circle. The latter algebraic exponential and series expression for sine and cosine are also fundamentally built up and predicated on the above equation which lays down the specification of what sines and cosines are.

The series expansion expressions and the algebraic exponential expression for sine, cosine, sinh and cosh are important and are good tools in certain applications but the above expressions do not define. However they do add in great measure to one's basic and fundamental understanding and appreciation of sine, cosine, sinh, cosh and $e^{j\theta}$.

EQUALITIES OR EQUIVALENCIES OF CIRCULAR AND HYPERBOLIC FUNCTIONS

It was stated on page 235 that as a valid, reasonable and tenable doctrine, one could select $e^{j\theta}$ and $e^{-j\theta}$ as vector operators. These operators would indicate counterclockwise and clockwise rotation in vector considerations of the problem at hand. In so doing, it would be rightfully assumed that the various rules and procedures of algebra and exponential techniques would prevail. In other words, the regularly accepted "machinery" of mathematics could be used validly to produce whatever results came forth. The rules and truths regarding j , j^2 , j^3 , etc. and the series expansion doctrines noted on page 235 would be applied also.

By exploring and maneuvering with these truths, one achieves some most interesting and important results regarding sine; cosine, sinh and cosh.

We know that $\sinh$ equals $\frac{e^x - e^{-x}}{2}$.

Substituting $j\theta$ for X , we write $\sinh j\theta = \frac{e^{j\theta} - e^{-j\theta}}{2}$.

As per previous pages, (235, 236), valid, reasonable and straightforward procedures developed the relations

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

$$\sin \theta = -j \left(\frac{e^{j\theta} - e^{-j\theta}}{2} \right)$$

It readily follows that one could by substitution and re-arrangements, write down four equalities between circular and hyperbolic functions.

$$1. j \sin \theta = \sinh j\theta = \frac{e^{j\theta} - e^{-j\theta}}{2}$$

$$2. \cos \theta = \cosh j\theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

$$3. \sin j\theta = j \sinh \theta = j \left(\frac{e^\theta - e^{-\theta}}{2} \right)$$

$$4. \cos j\theta = \cosh \theta = \frac{e^\theta + e^{-\theta}}{2}$$

It is to be emphasized that these equivalencies between circular and hyperbolic functions are the result of valid, reasonable and ordinary mathematical procedures with a special assist from the vector operator j .

From a strictly mathematical point of view, the only difference between circular and hyperbolic quantities, whether angle or function, is the factor j , the $\sqrt{-1}$. It is quite surprising and even thrilling to realize that the "wedding" of these two types of quantities is performed by j . To use a chemical analogy rather than a human relations analogy, j is the catalytic agent. In certain types of electrical circuit analysis problems, the above facts, truths and equalities serve as most useful tools.

The equivalencies discussed, the rotating vector doctrine and the vector operators $e^{j\omega t}$ and $e^{-j\omega t}$ enable the engineer and scientist to have an insight into and to get answers to many problems which otherwise might not have been so readily resolved.

One must study further to grasp and apply these relations to practical problems of electrical circuits.

A SIMILARITY

One knows basically and understandingly from fundamentals that the equation of the equilateral hyperbola is $X^2 - Y^2 = 1 = \cosh^2\theta - \sinh^2\theta$. Again one can call for an assist from j and do some manipulation. Since $j = \sqrt{-1}$ and $j^2 = -1$, one can write

$$\cosh^2\theta + (-1 \sinh^2\theta) = 1$$

$$\cosh^2\theta + j^2 \sinh^2\theta = 1$$

$$\cosh^2\theta + (j \sinh\theta)^2 = 1$$

One is reminded of the circular equation, $\cos^2\theta + \sin^2\theta = 1$.

CONCLUSION

The circular sine and circular cosine are recurrent, cyclic and periodic functions. They are good tools with which to work and apply to many physical and electrical phenomena which are periodic.

As said numerous times before in this treatise, they involve a constant distance namely the radius (R) in the denominator of the ratios they represent. Circular radians are involved and in a circle, the maximum number of radians per cycle is 2π .

The hyperbolic sine and hyperbolic cosine are not recurrent and periodic factors which increase and decrease in numerical magnitude. Sinh and cosh keep on increasing indefinitely in numerical magnitude as more and more hyperbolic radians are involved. The hyperbolic radians do not get much larger in sheer circular angular degrees magnitude to be sure even if five, ten, or a hundred hyperbolic radians are involved but they do keep on angularly increasing.

Circular radians are intrinsically measured as $\frac{\pi}{2}$, π , and 2π or parts thereof. The conventional 360° are validly useful. As per conventional concept of angle, all circular radians are equal. However, π and 360° have no involvement in hyperbolic radians. As per conventional concept of angle, two hradians are not twice one hradian.

An infinite number of hradians add up to only $\frac{\pi}{4}$ (45°) as per conventional and usual concept of an angle.

The numerical values of sine and cosine have, for their maximum amount, unity (1). At $\frac{\pi}{4}$ (45°) they have exactly the same value (.7071). Up to $\frac{\pi}{4}$ on the way to $\frac{\pi}{2}$ the cosine is greater than the sine; but after $\frac{\pi}{2}$ the sine is greater than the cosine. At all times $\sin^2 + \cos^2 = 1$.

The numerical values of sinh and cosh have no limit to their maximum value. Even though they approach near equality (201.7156) for cosh and 201.7132 for sinh) at six hradians, they can never, due to their intrinsic nature, be exactly equal. See page 85, Fig. 38. As previously given page 77 the cosh of 10 hradians is 11,103.232919 and the sinh is 11,103.232875. The cosh is ever greater than the sinh. There ever must be enough difference to comply with the equation $\cosh^2 - \sinh^2 = 1$.

The only similarity is at zero (0) circular radians and hradians. Here the cosine and cosh are both unity (1) and sine and sinh are both zero (0).

It has been interesting and satisfying to the writer to prepare this discussion. Vectors, expansion series and exponentials have been used interrelatedly to make a broad attack on circular and hyperbolic functions.

The approach and method of presentation has been the step by step technique. From the simple, basic and readily understood concepts of arithmetic, algebra, geometry and trigonometry, the rather formidable equations were logically evolved. In every case, the individual steps up the stairs of understanding had to make sense and hang together. As an end result, one gets a relationship expressed by a seemingly complicated equation. When this final equation is looked at as a whole, it may appear frightening and overwhelming.

However, if one has understood the simple truths and the basic step by step logic, one needs to have no fear. The Lincoln Memorial in Washington is an overwhelmingly beautiful architectural structure as a completed structure. However, all the elements and components comprising the end result are also individually exquisitely simple and beautiful.

During the preparation of this treatise, the writer gleaned much help from a number of books. One of the books which was particularly helpful was written by De Bray and is entitled "Exponentials Made Easy."

In the introduction, De Bray nicely expresses some thoughts worthy of quoting at the end of this chapter.

"Some time ago, the author came across a certain little book. Although he was supposed to know all about the things explained, he found great delight in reading it." The certain little book, later called "this truly delightful little book," was "Calculus Made Easy" by Sylvanus Thomson. De Bray continues

"In so doing, he re-learned several things he had forgotten and learned a few others he had not chanced to meet before."

The writer has the temerity to hope that the thought embodied in the second quote may apply to some oldster who reads and studies this homily. De Bray continues

"There are few branches of mathematics which seem more puzzling to beginners than the study of imaginaries and hyperbolics; indeed, many students who are no longer scared when they see $\frac{dy}{dx}$ or an integral sign confess that the appearance of i or j in a mathematical expression gives them a nerve-shattering shock and the sight of sinh or cosh is a signal for undignified retreat."

Again, the writer has the temerity to hope that the youngster and the oldster too who reads and studies this treatise may feel calm and confident when he meets the supposedly horrendous j , sinh, cosh, $e^{j\theta}$ and $e^{-j\theta}$, e^x and e^{-x} .

If readers learn and understand by the study of this treatise as much as the writer by its preparation, thinks he has re-learned about things he knew before and learned for the first time about some things he did not know before and got some ideas he did not have before, the preparation so far will have some measure of justification.

CHAPTER 13

Application of Sinh and Cosh

It is now in order with the background given in previous pages to directly consider the use of sinh and cosh in attenuation and allied problems in communication.

The manner of attenuation of power, bearing speech along a telephone circuit, has been discussed in the previous pages. Actually of course, the wires of the telephone circuit have served as guides to electromagnetic waves. Therefore, it is wave energy that attenuates in a manner as prescribed by the geometric retrogression doctrine. The wave travels with high velocity and constant velocity provided the circuit is a so called "smooth" circuit. A "smooth" circuit means a circuit along which at every point, the impedance is the same as it is at every other point. The impedance is therefore uniform along the entire circuit. If there are distances or "stretches" in the circuit where, due to changes in the electrical characteristics of the circuit, the impedance changes then there will be reflections of the wave energy. This simple doctrine of reflections at points where impedance changes applies also to sound waves and light waves.

ELECTRICAL LINES OF INFINITE LENGTH

If a telephone circuit is infinite in length and is smooth, then no reflection of wave power will take place. The power will continuously and uniformly decay as per the equation.

$$P = P_0 r^d = P_0 e^{-cd}$$

The "r" is a constant ratio loss between successive miles and is less than unity (1); c is logarithmic decrement constant characteristic of the particular circuit under consideration; the d is distance in miles.

At the end of the infinitely long line, the power approaches zero. Theoretically the power will never be zero however infinitely long the circuit is. The end of this hypothetical line may be "open", short circuited or may have a resistance of any value across it. An interesting and important aspect of such an infinite line is that however the line is terminated, the sending power supply will not know the difference.

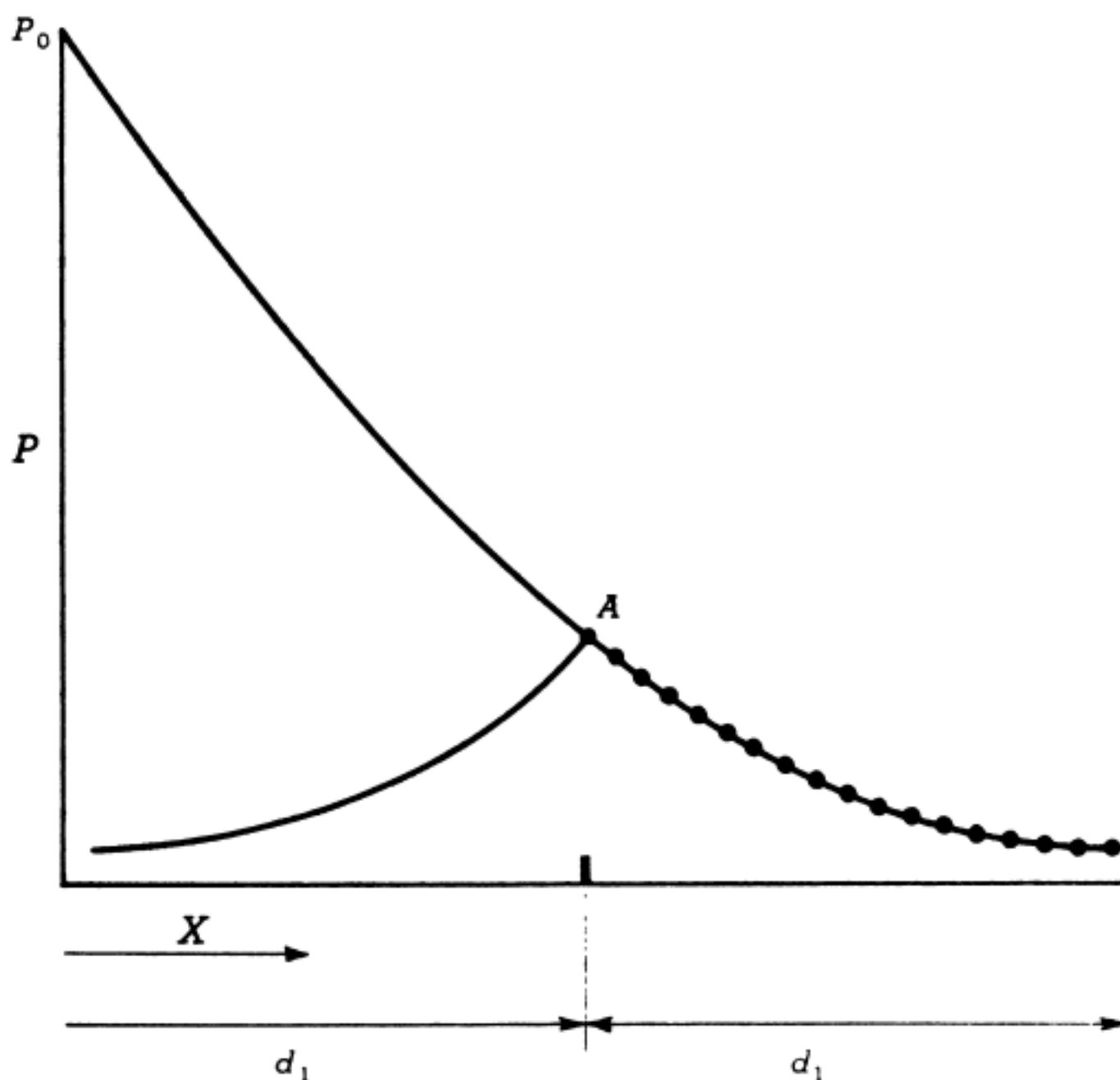
The infinite line will have a definite value of the characteristic or iterative impedance dependent upon its inductance, capacitance and resistance. The characteristic impedance of this infinitely long circuit will be the same however it is terminated. The power at every point along this infinite and smooth line will be given by the above equation.

Absolutely smooth and infinite lines do not exist. Carefully designed circuits of 3,000 or 5,000 miles in length behave practically as infinite and smooth lines.

FINITE LINES PROPERLY TERMINATED

A line of finite length can be made to simulate and perform as an infinite line, if the finite line is terminated by the characteristic impedance. The transmitter of the finite line will not know whether or not there is an actual infinite line beyond its own finite length provided there is connected across its terminal an impedance equal to its characteristic impedance if it were an infinite line. This doctrine propounds that the power sent out by the transmitter will be handed over completely to the termination, having the characteristic impedance, without reflection. See sketch page 249, Figure 95.

So, if one has a properly terminated finite line which simulates an infinite line, the equation for geometric attenuation applies. Knowing P_0 , " c " the attenuation characteristic or constant of the circuit, the power (P) at any distance can be readily calculated by use of tables of e^{-x} . There will be no reflections of wave energy and the basic equation will suffice. One needs no assistance from hyperbolic functions to get numerical values; all one needs is a table of e^{-x} values.



DECAY OF POWER ALONG AN ELECTRICAL
TRANSMISSION LINE •
REFLECTION AT POINT A

FIGURE 94

FINITE LINES IMPROPERLY TERMINATED: REFLECTIONS

Assume there is an impedance irregularity, giving rise to a change in velocity of the electromagnetic waves and in turn a reflection of those waves.

The discussion and arguments now to be presented can best be done in terms of the accompanying sketch, Figure 94.

One of the three basic and general attenuation equations is $P = P_0 e^{-\alpha d}$.

Assume there is a reflection at point A which is a distance of d_1 from starting point.

The power arriving at this point is $P_1 = P_0 e^{-\alpha d_1}$.

This wave power at d_1 is P_1 and reflection starts back on the line to the sending end of the circuit. This parcel of reflected power (P_1) attenuates in the same manner as the power P_0 did on its way out to the reflecting point at distance d_1 . This power P_1 at d_1 also attenuates in the same manner as it would have, had it not reflected but continued to go farther along the circuit. The power arriving back at the sending end from the reflection point is therefore $P_2 = P_1 e^{-\alpha d_1}$.

The distance out to reflection point is d_1 and the distance back to original starting point is of course d_1 too.

This parcel of reflected power P_2 arriving back at the original sending end is the power existing after a travel of $2d_1$ miles.

$$P_2 = P_0 e^{-2\alpha d_1}$$

TWO PARCELS OF POWER

If one now goes out on the circuit to any point at a distance of X from the sending end, there will be two parcels of power existent at that point. One parcel of power is part of the original power P namely $P_0 e^{-\alpha x}$. The other parcel of power is a bit of the reflected power from the point at a distance of d_1 .

One can now make a completely valid assumption and invoke a tenable doctrine. This assumption and doctrine says that the situation is as if there were two sources of power going out from the sending end. The first source of power is the original P and the second is P_2 .

As one goes out from sending end for a distance of X , the power P_0 decays and at distance of X , the power $P_{x1} = P_0 e^{-\alpha x}$.

The second source of power is P_2 which is $P_0 e^{-2\alpha d_1}$. As one goes from right to left on the lower curve from the reflecting point to the starting point, the reflected power P_1 at point d_1 decays as above equation $P_2 = P_1 e^{-\alpha d_1}$. However, one can imagine that one goes from left to right on this lower curve. As one goes from left to right on this curve, the initial power at starting point to point X increases according to the geometric progression doctrine. The equation applying then would be $P_{x2} = P_2 e^{+\alpha x}$; note the plus sign.

So in terms of the previous sketch, one can imagine that one parcel of power as it decays along the drooping curve from sending end combines with the second parcel of power as it increases along the arising curve from sending end.

In terms of an equation, one can write the above truths.

$$\begin{aligned} P_x &= P_{x1} + P_{x2} = P_0 e^{-\alpha x} + (P_0 e^{-2\alpha d_1}) e^{+\alpha x} \\ &= P_0 e^{-\alpha x} + P_2 e^{+\alpha x} \end{aligned}$$

Obviously the first parcel of power is always greater than the second parcel of power. As the distance X increases, the first parcel decreases in numerical value and the second parcel increases.

SOME ALGEBRAIC TRANSFORMATIONS AND SUBSTITUTIONS

We now have available in our thinking as previously presented in these pages, an abundance of rationalized concepts and doctrines on the one hand and a number of associated rationalized mathematical formulas on the other hand. We have exponential attenuation formulas and hyperbolic equations and equivalencies between them. By simple algebraic transformations we may be able to arrive at some important and significant ideas and formulas.

We know that

$$e^{-\alpha x} = \text{Cosh } \alpha x - \text{Sinh } \alpha x$$

and

$$e^{+cx} = \text{Cosh } cx + \text{Sinh } cx$$

In the equation for power P_z along an electrical circuit of finite length where reflections are present, the e^{-cx} and e^{+cx} appeared.

In this equation, $P_0 e^{-cx} + P_2 e^{+cx} = P_z$, we will substitute $\text{Cosh } cx - \text{Sinh } cx$ for e^{-cx} and $\text{Cosh } cx + \text{Sinh } cx$ for e^{+cx} .

By straightforward algebraic rearrangement we can readily write

$$\begin{aligned} P_z &= (P_0 + P_2) \text{Cosh } cx - (P_0 - P_2) \text{Sinh } cx \\ &= A \text{Cosh } cx - B \text{Sinh } cx \end{aligned}$$

AN OBSERVATION

The reader might validly observe that the above final equation involving sinh and cosh offers no advantage from the standpoint of time and labor saving over the equation $P_z = P_0 e^{-cx} + P_2 e^{+cx}$. In each case, one must look up two values in tables and then make a subtraction to get an answer. The observation is correct. The rebuttal is that it is always in order to have several approaches to every problem. It is of course remotely possible that one might have available sinh and cosh tables and not e^x and e^{-x} tables.

In this connection, it is also in order to refer to the previous curves which portrays decreasing alternating current on previous page 116, Figures 46, 47. The equation of the curve is $Y = 20 \sin \theta \cdot e^{-.2\theta} = 20 \sin \theta \cdot .818\theta$.

In some presentations of this equation, it is stated that it is easily evaluated by the use of a hyperbolic function table. This would involve the writing of $e^{-.2\theta}$ as $\cosh .2\theta - \sinh .2\theta$. The equation would then become $Y = 20 \sin \theta (\cosh .2\theta - \sinh .2\theta)$. The writer does not concur in this statement, since it is easier to evaluate $e^{-.2\theta}$ with varying values of θ from a book of tables than it is to find $\cosh .2\theta$ and $\sinh .2\theta$ and then subtract. It seems that the incorporation of sinh and cosh into the problem in this instance is not only unnecessary but also is actually rudely dragging them "into the act" by their heels.

CONCLUSION

The hyperbolic functions do then "get into the act" because of valid and tenable ideas on the one hand and because of mathematical relationships on the other hand.

As previously pointed out, hyperbolic functions are geometrically related to the hyperbola in a manner similar to the relation of circular functions to a circle.

However, it is not this geometric relation which is of significance to the electrical craftsman and engineer but the fact that algebraically and in complex or vector form, they represent two waves traveling in opposite directions. Furthermore they provide an easy and convenient way to get numerical values since their numerical values are available in tables.

The analysis of transmission line and wave filter problems and the role and use of hyperbolic functions calls for skill and competence. The practical problems involving these functions are not simple. The objective of this memorandum was not to discuss in detail their role and use but to develop a simple argument as to "how and why" they are involved in certain fields of endeavor.

As stated above, the matters discussed in this treatise and its predecessors are not as simple as a, b, c. Neither are they too difficult for just ordinary folks if given a good try.

The writer has the temerity to hope that after a good try and good study, the mere sight of e^x , e^{-x} , sinh and cosh will not give a "nerve-shattering shock to the reader when he encounters them". It is also hoped that the mere sight of these terms and the equations involving them will not be a "signal for undignified surrender and retreat".

The basic ideas, concepts, and doctrines and truths are after all simple and straightforward. Some ensembles may appear complex, difficult and perhaps a bit overwhelming. However, the simple components and basic elements can be worked at one at a time. Thereby a real problem and a practical situation is mastered and solved.

In the later pages, a discussion of a problem of somewhat different yet similar nature and involving e^x , e^{-x} , sinh and cosh will be presented.

CHAPTER 14

Infinitely Long Electrical Circuits: Use of Sinh and Cosh

Before continuing the interesting job of trying to illustrate why and how sinh and cosh enter the picture, it is in order to discuss a few of the basic principles and concepts of electrical circuits or lines. In so doing, one encounters the terms "infinite line" and "line extending to infinity." What is the import and meaning of these terms?

Infinity is not a definite and specific terminal at which the infinite line ends. Infinity is not a place or location but is, as has been observed, a state of mind. We know that there are celestial galaxies in the universe whose distances from the earth are measured in millions of light years. Such distances are beyond comprehension but they are not infinite. Infinity is far and away beyond those galaxies. So when one speaks of infinity, one must think of a magnitude that exceeds the greatest assignable magnitude however great that assigned magnitude may be.

As a practical affair, as far as electrical circuits are concerned, one can validly assume that a specific and finite length of the hypothetical infinite line is in effect a line of infinite length when a negligible percentage of the energy put in at the "sending end" reaches the designated distant point. In discussing negligibility in connection with damped waves and Q (quality of a circuit) page 165, a .1 % figure was assumed as the basis of comparison. However in an actual communication system, either radio or wire, the magnitude of the received power is very often much less than .1 % of the transmitter output power. A .1 % or .001 ratio would be a 30 decibel loss. In terms of decibels, losses in communication systems may 56 db or 60 db in a single part of a circuit. A 56 db loss means that only one part in 400,000 arrives at the receiving end. A 60 db loss means that only one part in a million arrives. In a radar system, the ratio of power at the radar receiver is 180 db down as compared to radar transmitter output power. The 180 db down means one-millionth of a millionth of a millionth. Even though the peak power output of the radar transmitter might be 100 kilowatts, the received reflected power would be one-tenth of one-millionth of one-millionth of a watt. This tiny received amount of power might appear negligible to a layman but it is not negligible to the electronic engineer who makes effective use of this wee bit of power.

Electrical noise interference is of course a controlling factor in establishing negligibility of signal power. One must always consider signal to noise ratio.

The geographical length of an electrical circuit that could be considered practically as "electrically infinite in length" would depend upon two factors.

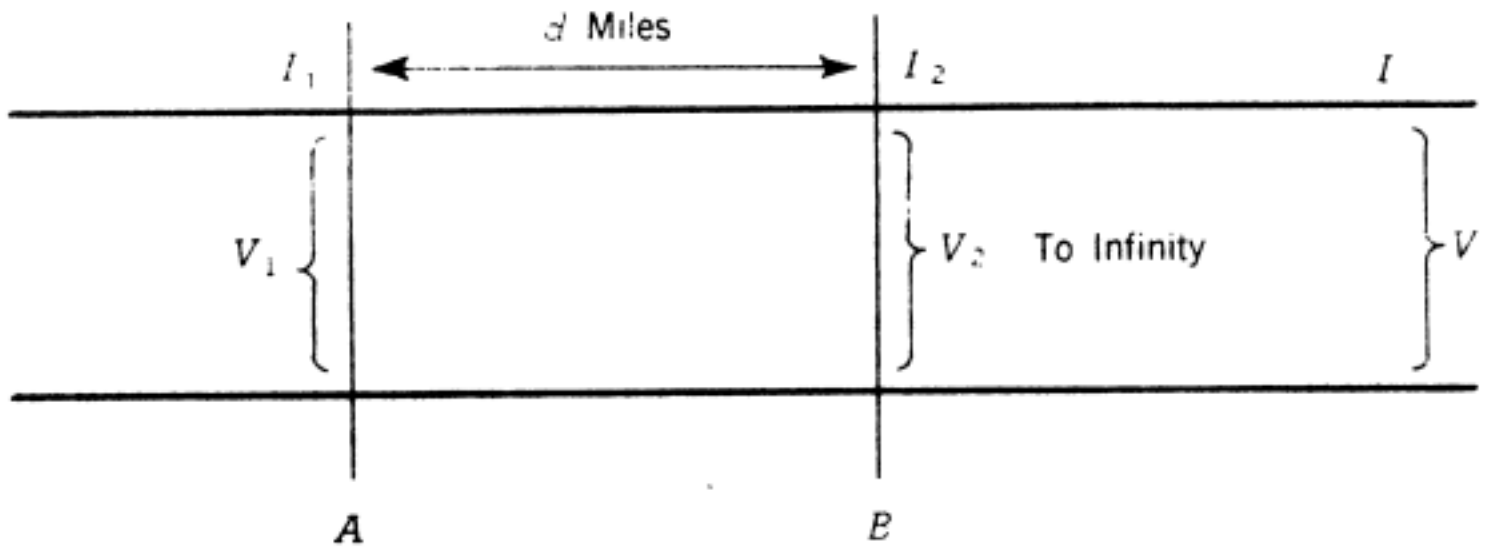
- (1) The rate of power loss or attenuations per unit length of the line in question, and
- (2) The percentage or the decibel loss of the input power that is considered negligible.

Some might regard a 1,000 mile circuit as a long line; others might regard a 3,000 or 5,000 mile circuit as practically equivalent to an infinite line but there always would be a difference of opinion.

STUDY OF FINITE LINES

It is desired to study and know the characteristics and performance of finite lines. As has been previously mentioned, the impedance presented at its sending end terminals by an infinite line is called characteristic impedance. To express the idea again, the term characteristic impedance of a specific type of line means the impedance presented or exhibited at the sending-end terminals by an infinite length of the type of line being considered. Also as has been previously stated, a line of finite length will accurately simulate, in characteristics and performance, an infinite line if the finite line is terminated by the characteristic impedance.

In the discussion and presentation now to be embarked upon, we will try to show how and why hyperbolic functions play a role in the study of electrical lines. It is not an absolutely indispensable role; it is not the lead-



SECTION OF INFINITE
ELECTRICAL TRANSMISSION LINE
FIGURE 95

ing role but it is a fine supporting role. To carry out this endeavor, we will consider, for illustrative purposes, the problem of the characteristics and performance of a circuit from a direct current point of view. This means that only the resistance components of the line will be considered.

We will assume that we have (but we do not have) an infinite uniform circuit composed of two copper wires at a fixed distance apart submerged in a very weak acid solution to provide a leakage or shunt resistance. For each unit length along the circuit there will be a series resistance equal to the ohmic resistance of the two wires and a shunt or leakage resistance equal to the resistance of the path between the two wires. If we imagine this hypothetical line consisting of many sections of equal and infinitesimally (there is that word again) small length, then we assume again that each one of the infinitely large number of sections would be identical electrically. Such a line would be called a "smooth line" in contrast to a "lumped" line.

One could not use voltmeters, ammeters or ohmmeters to determine the characteristic resistance (R_0) of the infinite line because one never has an infinite line.

However, let us say that according to the two criteria mentioned earlier in this discussion, agreement has been reached among several of those concerned as to the length of finite line that can be validly regarded as closely approximating and simulating an infinite line. The agreed upon length of this line is X and we have such a line stretching between two cities, A and B. Instead of the acid solution we will settle for real moist atmosphere along the circuit between the two cities in order to have some leakage or shunt resistance between the wires of the circuit.

Now that we have an actual circuit between two cities X miles apart, we can take measurements with voltmeter and ammeter to find the characteristics resistance, R_0 . It is to be observed and emphasized that if one took measurements with voltmeter and ammeter at sending end or at any desired distance from the ending end looking toward the terminal or far end, the same value R_0 for characteristic resistance would be obtained. This latter fact is a basic fact regarding an infinite line and also of our actual long distance circuit which has been agreed upon to very closely approximate and simulate an infinite line.

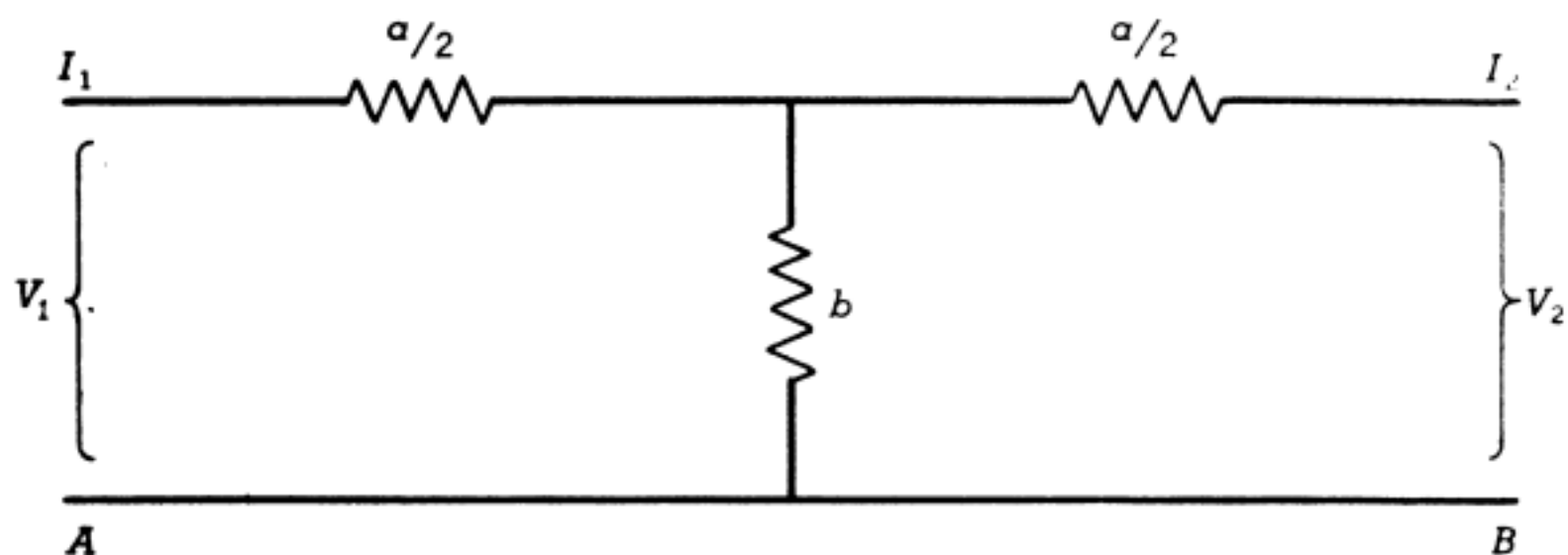
Since $R = \frac{V}{I}$ as per Ohm's Law, the fact that the same value for R_0 is obtained, wherever measurements are taken, means that the ratio of V to I is constant. Both V and I will attenuate in geometric retrogression fashion along the circuit but the ratio of V to I will remain the same: will remain R_0 .

It is desired, as said before, to study the characteristics and electrical performance of the X mile circuit but it is neither expedient nor practical to study it as such in its entirety. An X mile long circuit is a lot of circuit and stretches over hill, dale and rivers. So the engineer decides to construct a compact network or artificial line for study and experimentation in the laboratory, Figure 96. The specifications are that the compact network of series and shunt resistances shall precisely simulate electrically a finite length of " d " miles of the infinite line.

The problem is to find the numerical amount of the series and shunt resistances with which the compact network is to be made.

One appropriately might observe, since we know the series resistance per mile, why not make the lumped series resistance of the same amount that d miles of circuit would have. Then one would need to invoke no "fancy" calculations involving hyperbolic functions. The reason that this simple procedure could not be followed is that, in the actual circuit, a small and continuous quota of leakage or shunt resistance exists between every infinitesimal length of the series resistance of the actual circuit wires. Both the series resistances $\left(\frac{a}{2}\right)$ Figure 96 and the shunt resistance (b) of the artificial network will be physically small and "lumped" resistances. So to make sure that the " d " miles artificial network performs exactly the same as " d " miles of circuit, the values of the resistances in the compact network used for study and experimentation in the laboratory will need be found by allegedly "fancy" methods.

The argument, the doctrines, the analysis and the procedure whereby the formula by which the resistances can be calculated need not be "viewed with alarm". They will not be "fancy". Obviously attenuation, e^x , e^{-x} and maybe sinh and cosh will be encountered but "so what". We have a clear and explicit insight of them not only individually but also as they are intrinsically interrelated. We are their masters now and they are our talented servants.



ARTIFICIAL NETWORK
REPLACING d MILES
FIGURE 96

FINDING THE MILE RATE OF ATTENUATION OF CURRENT

As has been discussed, rationalized and emphasized many times, the decay or attenuation will take place according to the natural law of geometric retrogression. As one looks ahead to the job of finding the formulas for the resistances of the compact artificial network, Figure 96, it is readily anticipated that we should know the constant ratio " r " of the decaying currents at points in the circuit which are one mile apart. We write down therefore a basic truth.

$$\frac{I_1}{I_0} = \frac{I_2}{I_1} = \frac{I_3}{I_2} = \frac{I_4}{I_3} = \dots = r$$

The constant numerical value of " r " (ratio per mile) is less than unity (1). So, as has been illustrated many times before, to make the job easier one will invoke an exponential equation involving the logarithmic base " e ". The equations are

$$I_d = I_0 r^d = I_0 \left(\frac{1}{e^\alpha} \right)^d = I_0 e^{-\alpha d}$$

The procedure in the overall job is then for a man with meters to go out along the line of X length between the cities A and B wherever convenient and take measurements of current with an ammeter. It would likely need to be a milliammeter or a microammeter. It would be well to have a crew or team of test men at known distances along a goodly length of the line. On a synchronized time schedule, simultaneous readings could be made and repeated a number of times.

By this method, the " r " (ratio of decay per mile) could be determined. Knowing that $r = e^{-\alpha}$ one readily finds α from the table of e^{-x} . If the ratio of currents at a distance of two miles (r_2) happened to be made, the " r " (ratio of decay per mile) would of course not be $\frac{1}{2}$ of r_2 , but would be the square root of r_2 ; if 10 miles were involved, the " r " would be the 10th root of r_{10} .

To get " r " from actual measurements of current between any two points " d " miles apart, it would be easier and more convenient to use logarithms to find α .

$$I_d = I_0 e^{-\alpha d}$$

Taking logarithms of both sides of the equation

$$\begin{aligned} \log_e I_d &= \log_e I_0 - \log_e e^{-\alpha d} \\ &= \log_e I_0 - \alpha d \\ \alpha d &= \log_e I_0 - \log_e I_d \\ \alpha &= \frac{\log_e I_0 - \log_e I_d}{d} \end{aligned}$$

If " d " were 5, 15, 50 or 100 miles, the α or logarithmic decrement constant per mile is understandingly and easily calculated by use of the above idea and formula.

If according to the actual measurements of attenuating current along the circuit by a man with meters and use of the above formula, it happened that " α " came out to be .03, it would mean that " r ", which is equal to $e^{-.03}$, was .9704. An e^x and e^{-x} table provides the following:

$$e^{.03} = 1.0305; e^{-.03} = \frac{1}{e^{.03}} = \frac{1}{1.0305} = .9704$$

The current at a certain point would be 97.04 percent of the current at a point one mile nearer the input end of the circuit.

MOVING ALONG TO THE OBJECTIVE

In terms of the accompanying drawing Figure 96, let us do some simple analysis using some of the ideas already discussed regarding the relations of attenuation, e^x , e^{-x} , $\sinh$ and $\cosh$.

If one took readings of voltage and current at point A , the Ro as revealed by these readings would be $Ro = \frac{V_1}{I_1}$.

The Ro would be the so called characteristic resistance.

Now as per fundamental doctrine about attenuation and decay as it operates in many natural phenomena according to the geometric retrogression principle, the voltage and current decrease in this fashion as one goes out along the line portrayed in the sketch.

At point B , both the voltage and current are less than at point A . As for point A , $Ro = \frac{V_1}{I_1}$ so at point B , $Ro = \frac{V_2}{I_2}$.

All the way out on the line, there would be decreasing V 's and I 's but the ratio is always, Ro . Theoretically the current and voltage never become zero no matter how long the line. However far one goes toward the end of the theoretical infinitely long line, the $Ro = \frac{V}{I}$ relation prevails.

The current at point B can be found.

$I_2 = I_1 r^d$ where r is the loss ratio per mile. The " r " is numerically less than unity and " d " is the number of miles.

Also $I_2 = I_1 e^{-\alpha d}$ where α is the attenuation or logarithmic decrement constant characteristic of the particular circuit.

It is to be remembered that we know α as a characteristic of the circuit as determined by actual measurements by a man out on the circuit.

ARTIFICIAL NETWORK

In the place of the section of " d " miles of the infinite line, we now propose to substitute an artificial compact network for the " d " section. The basis of substitution is that this network will do nothing to change the electrical performance of the line as a whole. In other words, the transmitting end must not know that any changes have been made. Furthermore, the Ro is not to be affected. See Figure 96.

We will make this network as shown out of three resistances

$$a/2, a/2 \text{ and } b$$

Again it is to be stated that at point A , Ro must now be the same $\frac{V_1}{I_1}$ and at point B , Ro must remain $\frac{V_2}{I_2}$.

The problem is, what shall we use for the numerical values of resistances $\frac{a}{2}$ and b so that there shall be no differences of which the sending end might be aware. If we can find the proper values for the three resistances then instead of 100 mile section of line, we have a little box whose "insides" take the place of the 100 mile section of line. In this case, the " d " is assumed to be 100 miles.

It must be stated in the first place, before trying to find the desired concentrated resistances, " a " and " b ", that there is no attenuation or decay of current along them. In other words, the current on each side of the three resistances as such is the same. I_2 is now less than I_1 , only because some current is shunted or leaks through the resistance " b ".

OHM'S LAW AND KIRCHOFF'S LAW

Now we will apply Ohm's Law and its colleague Kirchhoff's Law to the network.

Kirchhoff's Law states the simple and almost obvious fact that the current through the resistance " b " of the network I_b is equal to $I_1 - I_2$. In other words, $I_1 = I_2 + I_b$ and $I_2 = I_1 - I_b$.

The voltage at A , namely V_1 which is to be exactly the same as it was before the network was substituted is given by Ohm's Law and Kirchhoff's Law.

$$V_1 = I_1 \cdot \frac{a}{2} + I_b \cdot b = I_1 \cdot \frac{a}{2} + (I_1 - I_2)b = \text{the sum of two voltage drops.}$$

As stated before $I_2 = I_1 e^{-\alpha d}$

Therefore $V_1 = I_1 \cdot \frac{a}{2} + (I_1 - I_1 e^{-\alpha d}) b$

In this last equation, all currents are I_1 's, which is of good advantage.

Now V_2 at point B is to be exactly the same as it was before artificial network substitution

$$\begin{aligned} V_2 &= I_1 \cdot b - I_2 \cdot \frac{a}{2} \\ &= (I_1 - I_2)b - I_2 \cdot \frac{a}{2} \\ &= \text{The difference of two voltages} \end{aligned}$$

In this equation, there are two places where I_2 occurs and one place where I_1 occurs. Could we get an equation with all I_1 's? Do we know I_1 in terms of I_2 ?

If instead of assuming that we go out farther and farther along the line to its end during which journey the current decays as per simple equation $I_2 = I_1 e^{-\alpha d} = I_1 \cdot \frac{1}{e^{\alpha d}}$, we now assume that we come in from the far end toward the sending end that is from right to left. Obviously if we do, the current would grow and geometrically increase.

According to above equation I_1 is greater than I_2 . So if we do come in from far end toward the sending end, we can readily write $I_1 = I_2 e^{+\alpha d}$

Of course I_1 was obtained readily by algebraic change in the above expression for I_2 .

The plus sign in above equation means a growth rather than a decay. This last equation says that I_1 at point A has geometrically progressed from its value I_2 at point B when one goes from right to left.

Now the equation for V_2 above was $V_2 = (I_1 - I_2)b - I_2 \cdot \frac{a}{2}$.

So now we can validly and understandingly write this equation with all I_2 's.

$$V_2 = (I_2 e^{\alpha d} - I_2)b - I_2 \cdot \frac{a}{2}.$$

We now have an equation for V_1 with the currents all I_1 's and an equation for V_2 with the current all I_2 's. Proceeding with some straightforward algebraic substitutions and transformations, we can readily write

$$\begin{aligned} Ro &= \frac{V_1}{I_1} = \frac{a}{2} + b - b e^{-\alpha d} \\ &= \frac{V_2}{I_2} = b e^{\alpha d} - b - \frac{a}{2} \end{aligned}$$

Adding these two latter equations $2 Ro = b (e^{\alpha d} - e^{-\alpha d})$

By some more simple algebraic transformations

$$b = \frac{Ro}{\frac{e^{\alpha d} - e^{-\alpha d}}{2}} = \frac{Ro}{\sinh \alpha d}$$

We now have derived an equation relating "ro" and "b". We know from previous measurements what Ro and α are numerically, so under the directives of the derived formula, now we can readily calculate "b" by reference to a hyperbolic sine table. It is to be recalled that "d" equals 100 miles.

AN ILLUSTRATIVE EXAMPLE

Assume that the characteristic resistance (Ro) for a particular circuit was found to be 300 ohms. A voltmeter and ammeter reading at circuit input provided this value.

Also assume that measurements by the crew of test men revealed that the current in amperes at a certain point was .9512. of the current at a point one mile nearer the input end. The ratio of attenuation or geometric retrogression per mile was .9512.

$$\frac{I_1}{I_0} = .9512 = r$$

The value of .9512 is assumed to be constant for every mile of the circuit.

If " r " = .9512, then $e^{-\alpha d}$ must equal .9512. As per table $e^{-.05} = .9512$. Also as per table, $e^{+.05} = 1.0513 = \frac{I_0}{I_1}$.

$$e^{-.05} = \frac{1}{e^{.05}} = \frac{1}{1.0513} = .9512 = \frac{I_1}{I_0}$$

The logarithmic decrement constant (α) for current of the illustrative circuit is .05 nepers; .05 times 100 miles = 5 = αd .

The shunt resistance (b) of the artificial network which is to be substituted for 100 miles of circuit is readily calculated

$$b = \frac{300}{\sinh \alpha d} = \frac{300}{\sinh 5} = \frac{300}{74.20} = 4.043 \text{ ohms}$$

HYPERBOLIC FUNCTIONS PLAY A PART

The denominator in the first part of the equation for " b " on preceding page, Fig. 96, is at a glance recognized as $\sinh$. We know α , we know d and we know Ro .

As a matter of sheer ease of calculation, we can find readily, with the help of $\sinh$ tables, the numerical value of the resistance " b ", out of which we wanted to make the artificial network which is to take the place of " d " miles of circuit.

Since we know αd which is the " x " in e^x and e^{-x} , we could look quite readily in the book of tables of e^x and e^{-x} . We could then subtract the two numbers; divide the answer by 2 to get the denominator. But why do all this arithmetic; it is easier and it is good sense to use a table with values for $\sinh$.

Some specialists and calculators have happily prepared tables of $\sinh$ and $\cosh$. All we have to do now is to use those convenient tables of numerical values for $\frac{e^x - e^{-x}}{2}$ and $\frac{e^x + e^{-x}}{2}$ the $\sinh x$ and $\cosh x$, respectively.

CALCULATION OF " $a/2$ ", FIGURE 96

Since we now know what numerical value the resistance " b " must be, we can go back to expression $\frac{V_1}{I_1} = Ro = \frac{a}{2} + b - b e^{-\alpha d}$ to easily calculate the numerical value of $a/2$. Therefore, the artificial network can now be built since numerical values of " b " and " $a/2$ " have been calculated by the help of simple, basic, logical and careful arguments and tables of hyperbolic functions.

Although the calculation of $a/2$, by substituting the known values of b and $b \cdot e^{-\alpha d}$ in the above equation, would not be burdensome, it would be well if one could get a formula which would permit easier calculation. It is therefore in order to do a bit of exploration and see if hyperbolic functions could be used again to save labor and time. Their use will not enhance our understanding of the general problem confronting us but their use might reduce the sheer labor of calculation. This exploration and the good results obtainable follow.

We know as a starting point $b =$

$$b = \frac{Ro}{\frac{e^{\alpha d} - e^{-\alpha d}}{2}} = \frac{Ro}{\sinh \alpha d}$$

$$\text{and } Ro = \frac{a}{2} + b - b e^{-\alpha d}$$

By substituting for "b" its value in the equation for Ro , one has by simple algebra

$$Ro = \frac{a}{2} + \frac{Ro}{\frac{e^{ad} - e^{-ad}}{2}} - \frac{Ro e^{-ad}}{\frac{e^{ad} - e^{-ad}}{2}}$$

$$\frac{a}{2} = Ro - \frac{Ro}{\frac{e^{ad} - e^{-ad}}{2}} + \frac{Ro \cdot e^{-ad}}{\frac{e^{ad} - e^{-ad}}{2}}$$

This equation may look pretty overwhelming but undaunted by e and negative exponents, one continues.

$$\begin{aligned} \frac{a}{2} &= Ro \left[1 - \frac{2}{e^{ad} - e^{-ad}} + \frac{2e^{-ad}}{e^{ad} - e^{-ad}} \right] \\ &= Ro \left[\frac{e^{ad} - e^{-ad} - 2 + 2e^{-ad}}{e^{ad} - e^{-ad}} \right] \\ &= Ro \frac{e^{ad} - 2 - e^{-ad}}{e^{ad} - e^{-ad}} \end{aligned}$$

At this point in the procedure, one does a bit of maneuvering. It occurs to one that e^{ad} is the square $e^{\frac{ad}{2}}$. This is true because $\left(e^{\frac{ad}{2}}\right)^2 = e^{ad}$.

So with this fact at one's disposal, one proceeds by substitution in the last expression.

$$\frac{a}{2} = Ro \left[\frac{\left(e^{\frac{ad}{2}}\right)^2 - 2 + \left(e^{-\frac{ad}{2}}\right)^2}{\left(e^{\frac{ad}{2}}\right)^2 - \left(e^{-\frac{ad}{2}}\right)^2} \right]$$

Furthermore one observes that the numerator looks like $(a - b)^2 = a^2 - 2ab + b^2$. Also one remembers that $e^{\frac{ad}{2}}$ times $e^{-\frac{ad}{2}} = e$ to the zeroth power (e^0) and $e^0 = 1$.

Also one observes that the denominator looks like $(a^2 - b^2)$ and that this is equal to $(a - b)(a + b)$.

So one can write the above as follows:

$$\begin{aligned} \frac{a}{2} &= Ro \left[\frac{\left(e^{\frac{ad}{2}}\right)^2 - 2e^{\frac{ad}{2}} \cdot e^{-\frac{ad}{2}} + \left(e^{-\frac{ad}{2}}\right)^2}{\left(e^{\frac{ad}{2}} + e^{-\frac{ad}{2}}\right)\left(e^{\frac{ad}{2}} - e^{-\frac{ad}{2}}\right)} \right] \\ \frac{a}{2} &= Ro \left[\frac{\left(e^{\frac{ad}{2}} - e^{-\frac{ad}{2}}\right)^2}{\left(e^{\frac{ad}{2}} + e^{-\frac{ad}{2}}\right)\left(e^{\frac{ad}{2}} - e^{-\frac{ad}{2}}\right)} \right] \\ &= Ro \left[\frac{e^{\frac{ad}{2}} - e^{-\frac{ad}{2}}}{e^{\frac{ad}{2}} + e^{-\frac{ad}{2}}} \right] \end{aligned}$$

It is now in order and valid to divide arbitrarily both numerator and denominator 2. The value of the ratio will be unaffected

$$\frac{a}{2} = Ro \left[\frac{\frac{e^{\frac{ad}{2}} - e^{-\frac{ad}{2}}}{2}}{\frac{e^{\frac{ad}{2}} + e^{-\frac{ad}{2}}}{2}} \right]$$

By algebraic transformations and substitutions in simple and logical steps, the numerator has been made $\sinh \frac{\alpha d}{2}$ and the denominator has been made $\cosh \frac{\alpha d}{2}$. Therefore

$$\frac{a}{2} = Ro \left[\frac{\sinh \frac{\alpha d}{2}}{\cosh \frac{\alpha d}{2}} \right] = Ro \tanh \frac{\alpha d}{2}$$

Now one can easily divide αd by 2 and then consult a table of $\tanh$. One will then have the numerical value for the resistance being sought in one "fell swoop."

Finally then

$$\frac{a}{2} = Ro \tanh \frac{\alpha d}{2}$$

Time and work will be save in calculation. The ever present attenuation in geometric retrogression fashion has been involved. The exponentials e^x and e^{-x} are expressions for geometric progression and retrogression. The e^x and e^{-x} have a numerical relation to $\sinh$ and $\cosh$. The hyperbolic functions get "into the act" because they can help make calculations easy; because they are a convenient and useful tool but not an indispensable tool.

It is smart and it is good practice to have and use good tools.

ANOTHER ILLUSTRATIVE EXAMPLE

As per previous illustrative example, the Ro was assumed to be 300 ohms. The logarithmic decrement constant (α) was found to be .05, on the assumption of " r " being .9512. The " b " was calculated to be 4.043 ohms

$$b = \frac{Ro}{\sinh \alpha d} = \frac{300}{\sinh 5} = \frac{300}{74.20} = 4.043 \text{ ohms}$$

It is now in order to determine the $\frac{a}{2}$ for the artificial network by means of the above beautiful and simple formula.

The $\frac{\alpha d}{2} = 2.5$; the $\tanh 2.5 = .9866$.

So $\frac{a}{2}$ one of the two series resistances of the so-called T network is 300 times .9866 = 295.98 ohms.

APPLICATION TO ALTERNATING CURRENT

The above equations and results are entirely valid for alternating current provided Ro if replaced by Zo , the characteristic impedance. Impedance involves inductance and capacitance as well as resistance and is a vector quantity. The attenuation constant α must be replaced by γ (gamma) which is the propagation constant.

The $\gamma = \alpha + j\beta$. The β (beta) is the phase or wave length constant. The β is related to the time required for the transmitted energy to traverse unit length of line.

A GENERAL PROBLEM

In many such cases as the one above discussed, one seeks to know numerical values. To aid in this job, one goes at the job with simple truths and logical step by step arguments. He arrives at equations which tell what the interrelations of his terms are. These terms in communication almost always involve attenuation and decay of power, current and voltage. Therefore the doctrines, principles, the equations and the mathematics bring e^x , e^{-x} , $\sinh X$ and $\cosh X$ into the picture. One does not need to look at these equations involving exponentials and hyperbolic functions as a "signal for undignified retreat". One need not need to "view them with alarm."

As a matter of fact, an exponential is a quarterback signal to specify a simple way to make a calculation of a simple geometrical retrogression existent in principle in many natural phenomena. Sinh X and cosh X are good tools to make easy calculations of numerical values. Some one kindly worked hard to prepare the extensive tables of their numerical values.

The carpenter uses a sharp saw, a sharp chisel and a hammer. He does not stop every time he uses one of these tools to ask how it was made.

The violin and piano artists skillfully play their instruments to the delight of the audience. Neither they nor the audience need to know in detail how a violin and a piano is constructed by the other artists who made them. The violin and piano players should and do know and understand a lot about their instruments as instruments. However, when they are playing, their job is to play. It is most likely that the instrument maker cannot play with great skill on the instrument he fashioned. The mathematician and the calculators commendably supply their "tools" to the engineer and the practical, skilled craftsman. The former make their contribution by developing the "tools"; they do not need to know just how the "tools" are used. The latter need not always stop to rationalize the "tools"; they can go ahead and use them.

The engineer and designer ought to know in an understanding way, the why and wherefore of exponentials and hyperbolic functions. But he does not need to always review his knowledge thereof, every time he develops an equation involving these quantities and every time he wants to do some calculations in a convenient and easy fashion. He just goes ahead and uses the tables and invokes a blessing on the folks who prepared the tables.

CONCLUSION

One final important thought is to be expressed. In using numerical values of hyperbolic functions as a convenient way to get numerical values of other desired quantities, one must not try to associate the electrical problem at hand with the geometric nature of a hyperbola as such. One must not seek to interpret and understand the electrical principles in terms of the sheer geometry of the hyperbola. Such interpretation and understanding is neither tenable nor possible. It is, as the radio announcer says, purely coincidental and a bit of good luck too that numerical values of sinh, cosh and tanh are fine "tools" with which to get an answer to electrical network problems.

It is a bit of good luck that even though cosh and sinh have no exponential nature per se, the sum of cosh x and sinh x is e^x and cosh x - sinh x is e^{-x} . See curve page 229, Fig. 90.

A previous thought merits repetition and emphasis.

The Napierian logarithmic base which is designated by the Greek letter epsilon or just plain (e), whose numerical value is 2.7128, need not be "viewed with alarm" either. As said before " e " is sometimes called a natural number.

The e^x and e^{-x} need not frighten anybody. Neither the e , e^x nor e^{-x} have any magic properties and as such have no connotation per se especially applicable to electrical problems. As in a former statement about the hyperbola, one must not try to associate, interpret and understand electrical phenomena and principles with these latter numerical "tools" as such. These "tools" are just that; no more and no less. They enable one to get numerical values conveniently when one does understand about the decay and geometric retrogression doctrines which do seem to naturally enter into many physical and electrical phenomena.

The really important task for the scientist, engineer and craftsman is to thoroughly grasp and understand basic principles of electrical phenomena, techniques and instrumentation on one hand and three distinctively electrical doctrines or laws on the other hand. These three basic laws are Ohm's Law, Joule's Law (the electrical energy law) and Kirchoff's Laws. A certain amount of facility in mathematics is of course an added requirement.

THE "WHY" OF HYPERBOLIC FUNCTIONS

After all, hyperbolic functions serve only as calculators where attenuation and decay is concerned. It is to be remembered that they are not indispensable. The crux of the situation centers around the simple and basic fact and doctrine of geometric retrogression. The retrogression can be expressed mathematically in at least 3 ways; a fraction with an exponent, r^x and e^{-cx} (.368^{cx}) and 10^{-kx} (.1^{kx}). The three equations are exponential

equations. e^x and e^{-x} can be evaluated numerically in terms of $\sinh X$ and $\cosh X$. Therefore, hyperbolic functions are convenient numerical tools to get answers in electrical communications because decay and attenuation is ever present as electric waves go over a telephone line, through a network or a filter of a wire or radio system.

The circular sine as a mathematical function has a direct geometric association with the projection of the position of a point rotating in a circle on a vertical axis. Furthermore, the sine function has an interpretable and physical relation to the voltage produced in a conductor of an armature rotating in a uniform magnetic field. The sine function and its numerical values does have a direct relation with a physical and an electrical phenomenon.

On the other hand, hyperbolic sines do not have any interpretable or geometric association as such with a physical or electrical phenomenon. As previously pointed out, the numerical values of $\sinh$ and $\cosh$ are related numerically to e^x and e^{-x} . These fortunate numerical relations are purely coincidental. The phenomena to whose quantitative relations $\sinh$ and $\cosh$ provide an "assist" involve geometric progression and geometric retrogression or attenuation. The e^x and e^{-x} do relate intrinsically to geometric progression and retrogression. The numerical values of $\sinh$ and $\cosh$ are, as stated above, so related to e^x and e^{-x} that they nicely serve to get numerical answers for the $\sinh$ and $\cosh$.

Again, for sake of emphasis, it is to be stated that the equilateral hyperbola, as far as its sheer nature is concerned, does not have any direct physical or geometric qualities to assist in interpreting and understanding of attenuation problems in electrical lines and networks.

In considering the role and use of $\sinh$ and $\cosh$ in electrical communication phenomena and equations, it is important and basically significant to appreciate the fact that, even though the geometry of the hyperbola and the nature of $\sinh$ and $\cosh$ per se are not intrinsically related to problems in this field, the difference of $\cosh$ and $\sinh$ provide a geometric retrogression and the sum of $\cosh$ and $\sinh$ embody a geometric progression. Since attenuation, occurring in geometric retrogression fashion, does enter into so many natural phenomena, the relation $\cosh - \sinh = e^{-x}$ is a most important and valuable one, Fig. 90, page 229.

EPILOGUE

The author has, for many years in his long life, studied and thought about the matters presented herein. Before he started writing this book in which he has used the "fugue" technique, he thought he knew and understood quite a bit about the whole "deal". He is happy to say that he has learned a lot in the preparation of this book. He hopes that the reader will not only have learned a lot, too, but will also have enjoyed his learning as much as the writer has.

In the many treatises and textbooks which the writer has read and studied for many years, he has not found many of the important fundamental and interesting ideas and facts that he evolved while preparing the book. He has tried to express clearly and simply his present understanding. Of one thing he is sure, there is always the possibility of knowing more and more about the things one thinks he knows something about. As a continuing process, one increases his competence and skill by always trying to learn more about those matters relating to the particular job at hand.

The easy essentials and basic fundamentals have been the objective in this treatise. In any endeavor, be it teaching or practicing, one must be well versed in the essentials and fundamentals. After and beyond these, one may go to the more advanced. One cannot solve the more advanced and difficult by merely knowing the fundamentals. In terms of the well mastered fundamentals, one must solve more complicated problems. The more complicated problems are a bit "tougher," but one does not need to be a "super duper" shark to understand and solve them. To be sure, some increased measure of skill, competence and understanding must be had to do the job. "When the going gets tough, the tough get going".

To those tough readers who have a desire and a disposition to continue further and study advanced textbooks and treatises, the writer says "Splendid!" It will be exciting and profitable, and not an overwhelmingly difficult adventure. To such readers, the author extends best wishes and a hearty "bon voyage."

Index

- Absorption and Attenuation of Light, 54-58, 114
- Acceleration, Mechanical and Electrical, 46, 141, 149, 177, 183
- Alternating Current, 115, 116, 149, 153, 154, 159-161
- Ampere per Second, Carty 146, 149, 152, 177, 183
- Amplifier Analogy, Mechanical, 123
- Amplitude, Tuning Forks, A. C., 114, 115, 149, 177, 183
- Application of Sinh, Cosh, 244
- Arc, Circular, Hyperbolic, 70, 79, 80
- Areas, Circular, Hyperbolic, 21, 26, 42, 62, 70, 73, 77, 79, 218
- Arithmetical Progression, 46, 98
- Asymptotes, 66, 80
- Attenuation, 2-5, 96, 97, 102, 114
- Atmospheric Air Pressure Gradient, Exponential, 114
- Bels, 134, 135, 136
- Cable, Telephone, 124, 130, 131
- Capacitance, Farad, 139, 154-158, 175, 178, 183, 191
- Carty, Ampere per Second, 146, 149, 152, 177, 188
- Charging; Capacitor, Inductor, 147, 155-158, 193, 205
- Circle, 21, 23, 29, 33, 34, 37
- Circuit, Electrical, 96, 113, 115, 130, 131, 134, 244, 249
- Circular Pendulum, 38
- Compliance, Electrical, Mechanical, 155
- Compound Interest, Spending, 100-104
- Conic Sections, 7, 87, 90
- Conjugate Hyperbolas, 64, 66, 67
- Constants, Logarithmic Decrement, 4, 119, 166
- Cosh, Definition, 69, 76, 77, 79, 84
- Cosh, Series Expansion, 226, 227
- Cosine, Circular, 21, 25, 26, 36, 41, 83, 149
- Cosine, Hyperbolic, 69, 74, 75, 78, 149, 227, 228, 244
- Damped Wave Train, 117, 162-165, 238-240
- Decibels, 107, 127, 131, 134, 135, 244
- Decrement, Logarithmic, 166
- Definition of Sinh and Cosh, 69, 74, 75, 78
- DeMoivre's Theorem, 234
- Directrix, 32, 41, 43
- Effective Value of A.C., 154
- Efficiency in R, L and C circuits, 214, 219, 222, 224
- Electric Field, 189, 155, 160, 180-219
- Electrical Circuits, 96, 113, 115, 130, 131, 134, 244, 247
- Epilogue, 259
- Ellipse, 7-20, 30, 32, 35, 39, 49, 50, 87
- Elliptical Pendulum, 29, 30, 37, 39
- Energy in R, L and C circuits, 186, 202, 205, 215
- Energizing Capacitor, Inductor, 147, 155, 158
- Epsilon, ϵ , Numerical value, 102, 130
- Equation of Circle, 24, 25
- Equation of Musical Scale, 99
- Equations of Transients, 176-178
- Equivalencies of Circular and Hyperbolic Functions, 241, 242
- Exponential Sine, Cosine, Sinh, Cosh, 83, 84, 222, 226, 227, 230
- Exponentials, Positive, Negative, 2, 96, 99, 108, 114, 120, 175, 225
- e , Napierian Logarithmic Base, 2-6, 101, 102, 130, 133
- e^x , e^{-x} , 101, 102, 130, 138, 230, 234, 258
- Expansion, Series, 81, 102, 180, 225, 226
- Farad, Static and Dynamic Definition, 158
- Finite Electric Lines, 246
- Frequency, Kelvin Formula, 60, 131, 162-165
- Geometric Progression, 2-6, 42, 96, 98, 104, 113, 120
- Half-life, Radium, 110-112
- Harmonic Motion, 26, 29, 37, 38, 44, 48, 50
- Henry, 140, 141, 145, 146
- Horn, Exponential, 108, 119
- Hyperbola, 7, 60, 62, 66, 87
- Hyperbolic Areas, 73, 77, 78
- Hyperbolics: Sinh, Cosh, Tanh, 69, 70, 73, 76, 80, 86, 96, 132
- i , 234, 236, 241
- Inductance, Henry, 139, 140, 141, 144-147, 160, 175, 180
- Infinite Electrical Lines, 244
- Interest, Compound, 100-102, 104
- j , $\sqrt{-1}$, i , 234, 236, 241
- Kelvin, Fundamental Equation of Oscillation, 162-175
- Kettering, Acceleration, 145, 146
- Kirchoff's Laws, 141, 162, 177, 178, 253
- Law, Kirchoff's, 141, 162, 177, 178, 253
- Law, Natural Geometric Progression and Retrogression, 96, 100, 113, 114, 120, 130, 252
- Law, Newton's Cooling, 114
- Light Absorption, Reflection, 54-58, 114
- Lissajou, Curves, 49, 50, 54
- Lines of Infinite Length, 244
- Logarithmic Decrement, 166
- Logarithms, 2-10, 99-101, 127-133
- Magnetic Field, 145, 180, 219
- Mathematics, General Use, 2-6, 257
- McLaurin Series, 81, 102, 180, 225, 226
- Mechanical Capacitance, 155
- Mechanics, Compliance, 120, 155
- Mirror, Parabolic, 54-58
- Musical Scale, Exponential, 98, 104-108
- Napierian Logarithmic Base, e , 102, 130, 133, 180
- Napierian Logarithms, 101, 102, 130-133, 180

- Negative Exponents, 2-7, 114, 134, 175
 Nepers, 13-138
 Network, Electrical, 244-248, 250, 254-257
 Newton's Law of Cooling, Exponential, 114
 Number of Oscillations in Wave Train, 118-119, 166-168
 Numerical Values of Sine, Cosine, Sinh, Cosh, 81-85, 226
 Ohm's Law, 253
 Operator, Vector, 233-238
 Organ Pipes, Exponential, 98, 104-107
 Oscillating Circuits, 139, 154, 155, 162-163, 172, 173, 214
 Oscillations, Number of, 118-119, 166-168
 Oscilloscope Curves, 50-53
 Parabola, 7, 40-59, 87-95
 Pendulum, 29
 Pi, Numerical Value by Series Expansion, 86, 180
 Planimeter, Area, 218
 Power Driven Winch, 119-123
 Power in Electric Circuit, 130-134, 211-214, 246
 Power in R, L and C Circuits, 211-217
 Progression, Arithmetic, Geometric, 48, 49, 96
 Pythagorean Theorem, 21, 23
 Quality of an Oscillatory Circuit, 165-171
 Radians, Circular, Hyperbolic, 69-80, 149-151, 242
 Radium, Half-life, Exponential, 111-112
 Reactance, 141-148, 154, 160
 Rectangular Hyperbola, 66-68
 Reflection, 54-59, 244, 245
 Relation of e^x , e^{-x} , Sinh, Cosh, 225-229
 Resistance, Ohm, 139, 140, 156, 161, 183, 214, 255, 257
 Resonance, 154-161
 Rotation, Circular and Elliptical, 30, 31, 37
 Rotating Vectors, 233-239
 Series Expansion; Taylor, McLaurin, 81-83, 102, 180, 225, 226
 Simple Harmonic Motion, 25-31, 37, 48-53
 Simultaneous Motion at Right Angles, 25-31, 38, 48-53
 Sine, Circular, 21, 26, 36, 81, 149-151
 Sine, Hyperbolic, 69-78, 227-229, 245-249
 sinh, Definition, 69-79
 Sinh, Expansion Series, 83-84
 Spending, Compound, 103
 Springs, Coiled, 154
 Submarine Cables, 219
 Taylor Series, 81-83, 102, 180, 225, 226
 Telephone Cable, 119-132
 Telephone Power Curve, 124-134
 Termination of Lines, 246
 Time constant, 171-172, 175, 183-221
 Transients in R, L and C Circuits, 175-224
 Triangles, Right, 21, 22
 Trigonometry, 21-24, 42, 71, 72
 Tuning Forks, Amplitude, Attenuation, 117
 Vacuum Tube Amplifier analogy, 123
 Vectors, 225-238
 Velocity of Electromagnetic Waves, 60, 173, 174
 Wave Length, Electromagnetic, 60, 173, 174
 Wheelbarrow Analogy to Inductance, 141-146
 Winch, Mechanical, Exponential, 119-123